The effects of two interventions on teaching quality and student outcome

Azkiyah, Siti Nurul

The Effects of Two Interventions on Teaching Quality and Student Outcome

A comparison of education standards and education standards combined with a teacher improvement program in Indonesia

Siti Nurul Azkiyah

university of groningen / faculty of behavioural and social sciences

ico Interuniversitair Centrum voor Onderwijswetenschappen
Interuniversity Center for Educational Sciences

Cover Design : Aini Kholid

Layout : Jay Jauhari

RIJKSUNIVERSITEIT GRONINGEN

The Effects of Two Interventions on Teaching Quality and Student Outcome

A comparison of education standards and education standards combined with a teacher improvement program in Indonesia

Proefschrift

ter verkrijging van het doctoraat in de

Gedrags – en Maatschappijwetenschappen

aan de Rijksuniversiteit Groningen

op gezag van de

Rector Magnificus, dr. E. Sterken,

in het openbaar te verdedigen op

donderdag 23 mei 2013

om 16.15 uur

door

Siti Nurul Azkiyah
geboren op 11 mei 1976
te Boyolali, Indonesie

Promotores:	Prof. dr. M.P.C. van der Werf Prof. dr. B.P.M. Creemers
Copromotor:	Dr. S. Doolaard
Beoordelingscommissie:	Prof. dr. L.E. Kyriakides Prof. dr. P.J.C. Sleegers Prof. dr. T. Wubbels

ISBN: 978-90-367-6237-3 (Paper version)

978-90-367-6260-1 (Electronic version)

Acknowledgment

I remember in my childhood I liked to learn when I thought my teachers made me understand what I learned, challenged me with what I should learn more, and helped me realized the importance of what I learned. Without knowing EER, in addition to my parents, since my childhood I have thought that teachers have been the most influential actors in my educational journey. This thought has made me decide to work on teacher and teaching when I got the opportunity to pursue a PhD study. Moreover, when I had the intention to do this study, teachers in Indonesia have been reported to have problems dealing with the newly launched education standards, based on which teachers are required to develop their own curriculum.

Yet, my PhD project would not be an experimental and longitudinal, a quality work that I have never previously imagined myself, without the presence of my supervisors. I learned a lot throughout the stages of the research, from the design, the pilot studies, the communication to schools and teachers, the implementation of the interventions, the measurement as well the data analysis and the interpretation. I came here to Groningen with a very limited knowledge of statistics and now not only I have analyzed my own data with multilevel modeling analysis but also I have been assisting analyzing a research project at the Indonesian Ministry of Education using multilevel analysis while I went back to Indonesia to wait for the PhD defense.

In the first place, I would like to deeply thank and extend my appreciation to Prof. Greetje Van der Werf, Prof. Bert Creemers, and Dr. Simone Doolaard for their thorough supervision, support and patience throughout the PhD period. Ibu Greetje, you are both a great promoter and manager. Thank you very much your clear and detailed feedback and your availability almost all the time I knocked on your door. I highly appreciate your dedication. When you could not attend our meeting, you either gave your written feedback or set up another meeting for only both of us. I do remember our meeting on the 2nd January 2012, in which many people were enjoying their New Year holidays. I recognize that you also worked during the weekends for the drafts of my thesis. Pak Bert, without your name being recognized by a friend of mine working in the Indonesian Ministry of Education, I might not end up doing my PhD at the University of Groningen. Thank you very much for your enthusiast response when I sent my PhD proposal. You are the first person to open the door for me to pursue my PhD at the University of Groningen, which I am proud of.

Simone, thank you very much for taking care of my administrative procedures when I came to the university at the first time and for recognizing “weird things” in the drafts of my thesis. You are the only supervisor that somehow make me think that I need to learn Dutch, *maar ik spreek Nederlands basis en betje en zo ik alleen met U zeggen hoe gaat het en tot ziens*. I also have to thank to Dr. Henk Guldemond for his invaluable assistance in analyzing my data, from which I have learned a lot. To Jessica Bakker, thank you very much for reading and editing this thesis. You did a great job. Finally, I would like to thank to the external examiners of the thesis: Prof. dr. L.E. Kyriakides, Prof. dr. P.J.C. Sleeger, and Prof. dr. T. Wubbels for approving this thesis.

Doing PhD in Groningen is a great experience, both academically and personally. Therefore, I would like to thank to the Training of Indonesia Young Leaders (TIYL) Program of the University of Leiden, which made it possible for me to do a PhD study. Prof. Wim Stokhof, Dr. Nico Captain, and Marise Van Amersfoot, thank you very much for your assistance. Pak Nico, thank you for allowing me to find another supervisor who suits my topic rather than changing my topic to suit the *existing* supervisor. My thanks also go Femke Groeneveld and Prof. dr. L. Buskens for their assistance during the very last stage of my PhD period. I also have to thank to Groningen Institute for Educational Research (GION), which has provided additional budget for important activities during the PhD study that were not covered by the University of Leiden. Ibu Greetje, I have to thank you again for this help.

To people in GION, thank you very much for all your support. Ridwan, Magda, Cata, Mayra, Coby, Mechteld, Anouk, Hanke, Willem, Anneke, Annemieke, Aboma, Wondimu, Marian and all that I cannot mention, for sharing both the ups and downs of being a PhD student and some strategies to finish the PhD study. Especially to Ridwan and Mayra, also thank you for being the paranymphs of my defense. To Sonja, thank you for your warm assistance and for keeping the Christmas gift when I was away. To Vera, *dank U wel voor uw helpen met praktische behoeften en mijn Nederlands*. To Djurre, thank you for your great help in cleaning my data. I have to mention that I am also grateful to the *laskar* of TIYL, Pak Din, Kang Kus, Agung, Mas Nur, Pak Amiq, Uda Yasrul, Aa Hilman for the togetherness that we shared especially at the very beginning of our study. Also people in my house, who have been very helpful, I owe a lot. Tante Pantja and Om Bas, you have saved my time by cooking for us more than twice a week and by taking care of Jingga after school

hours especially on Mondays and Fridays. Nurul, Mila, Ari, Mba Puri, you all have been great companions for me and Jingga. Then I have to thank to Indonesian students at the University of Groningen, Mba Ida, Mba Lia, Mas Yayok, Sistha, Mas Adhi, Mas Henki, Mba Erna, Mba Yuni, Astri, Bang Amir, Winarto and all that is too long to mention. I would like to re-mention some people who have been very helpful in bringing some of my paper data that must be checked due to a serious problem I had. Astri, Hilman, Pak Amiq, without your willingness to carry my data in your baggage, the data would not arrive as soon as I expected and I would have spent a lot of money to send them to Groningen. Furthermore, I do appreciate schools, teachers, students who have been willing to participate in this study and the big family of UIN Jakarta. Aida, Didin, Iim, Irman, Muji, thank you for collecting the data for this research. Ibu Itje, thank you for your inspiring talk and for sharing your great knowledge and experience during the monthly workshops for experimental group two in this study. Pak Nasrun, Pak Syauki, Pak Atiq, Bu Nida, Bu Neneng, thank you for your support. Pak Ahmadi and Mba Halimah in *rektorat*, thank you for your important assistance. My teachers in IKIP Malang (now State University of Malang), Bu Mimil, Pak Yudi, thanks for your support. I also thank my friends who have been willing to discuss some issues related to my research: Pak Bambang, Mas Nanang, Ibu Manik, Puti, Arif, and Mas Yanuar. To Anis and Eva, thank you for the laughs and jokes we shared during our shopping time when I came back to Jakarta for the field research. Finally, to Jay, thank you for your great work for the layout of this thesis.

Last but not least, I would like to thank to my family. To my parents, thank you for your endless pray and support. I thank my father who passed away before seeing this achievement. I thank my mother who has taught me to be an independent woman. To my brothers and sister in low, Arif, Iim, Udin, and Dewi, thank you for your support and for taking care of Jingga when I was away during the first year of my PhD study. Most importantly, I would like to sincerely thank to my beloved husband, Kelik, for giving me “his heart” to be away for such a long time and for his patience, trust, and love. For some moment, you functioned as both the father and the mother for Jingga and for another moment you were left with “emptiness” as both I and Jingga were in Groningen. Moreover, you also took care of my mom. Specifically to my daughter, Jingga, this work is dedicated to you and your future. Thank you for being so patience to wait for my promise to give a younger sister or brother after I complete this work.

Contents

CHAPTER 1

Introduction

INTRODUCTION

1.1 Research Background

Indonesia has been facing various educational problems, especially low student achievement, as indicated by both national and international measurements (e.g. Nurweni & Read, 1999; Jazadi, 2003; Kompas, 2010; Kompasiana, 2011; Mohandas, 2004; Setiogi, 2003; Tilaar, 1992). This problem is not only common in developing countries, like Indonesia. A number of research studies in the US for instance, observed both low student achievement and a continuing gap between school performance and socio-economic status across racial and ethnic groups (e.g. Education Commission of the States, 2002; National Commission for Excellence in Education, 1983; US Department of Education, 2008). In Indonesia, the government has adopted a number of strategies to improve the quality of education. A fairly recent example is the 'education standards', as formally established in 2005. This new policy is in line with the wide use of education standards in America (Dowson, McInerney, & van Etten, 2007; Schmidt, Houang, & Shakrani, 2009), in Europe (Neumann, Fischer, & Kauertz, 2010), in Australia (Donnelly, 2007), and in Asia (Choi, de Vries, & Kim, 2009).

It has been argued that education standards can serve as a guideline in improving teacher instruction by – first of all - providing a clear explanation of the expectations about what students, regardless of their backgrounds, are expected to accomplish (Marzano, 1998; Marzano & Kendall, 1996a; Ravitch, 1995). There is a strong belief that based on these standards teachers will adapt their ways of instruction to develop a more focused and improved approach to teaching and learning, which addresses the needs of all students to achieve the same minimum goals. Furthermore, all teachers of a particular subject are expected to foster the same expectations, while low-achieving students should not be used as an excuse to justify lower expectations. Based on these arguments, the education standards have been considered to increase both the excellence and equity in education (e.g. Buttram & Waters, 1997; McClure, 2005). In addition, education standards are also regarded to be promoting accountability, according to which schools have to report their results and bear the consequences, if necessary (Baines & Stanley, 2006; Chambers & Dean, 2000; Schmitdt, et al. 2009).

However, both in Indonesia and in other countries, it has until now remained unclear whether the claims of educational excellence and equity have already been fulfilled. The number of studies, especially experimental ones, which have investigated the effects of education standards, is limited. Especially in the US the existing studies state that the effects of education standards are still inconclusive, yet the bottom line is that the disadvantaged groups have simply remained to be left behind (Lauer, Snow, Martin-Glen, Van Buhler, Stoutemyer, & Snow-Renner, 2005; National Research Council (NCR), 2001; Thompson, 2009). Germany has a similar story. Here the effect of the education standards was evident only in the case of high achieving students (Neumann, et al. 2010).

These results are likely to be related to those of other research focused on problems with respect to the standards movement, for example difficulties associated with the standards documents, which are supposed to guide the instruction activities in the classroom. These documents have been found to be too general, unclear, vague and ambiguous, while they lack a solid grounding (Choi, et al. 2009; Gandall, 1996; Finn, Petrili, and Vanourek, 1998). Another important problem applies to the preparation and implementation of the standards, to which too little attention is paid, thereby leaving teachers in the dark. Also in Indonesia, the standards documents are considered as too broad and general, while no clear guidance is provided in how to translate their content into clearly formulated curriculum directives. In sum, apart from the fact that the provision of training and other types of assistance has been minimal, the standards documentation has only caused confusion among the teachers (e.g. Hanafie, 2007; Lampung Post 2006).

Thus, formulating the expectations about what should be expected of students' performance does not necessarily tell us how education should be improved. Educational improvement, however, is important for various reasons, such as quality enhancement and maintenance. It is logical to assume that policy makers, who are expected to help schools maintain and/or improve their quality, will want to know more about the empirically validated strategies of improvement. Moreover, the call for educational improvement is not merely based on the specific targets defined in the education standards. In this context, it is also highly relevant to look at the results of the Educational Effectiveness Research (EER), which provides both theory-driven and evidence-based information on factors related to student outcome.

After controlling for the various background characteristics, previous EER studies established that in the improvement of education the teacher is one of the most influential factors (Creemers, 1994; Darling-Hammond, 1997; Doolaard, 1999; Fullan, 2001; Harris, 2002; Harris & Muijs, 2005; Marzano, 2007; OECD, 1994; Pilot, 2007; Van der Werf, Creemers, De Jong, & Klaver, 2000). Therefore, it is crucial to prioritize improvement at the teacher level, a measure which simultaneously addresses the problems of the preparation and implementation stages of the standards movement.

This study has aimed to test the effectiveness of both the education standards on their own and in combination with a teacher improvement program. The standards had to be clear, specific, and understandable, while the teacher development program was based on the empirical findings on teacher factors related to student outcome, as provided by EER. In order to meet our research objectives, two interventions were organized, for which two supporting documents were developed. The first, called the elaborated standards document, was a response to the aforementioned problems associated with the original standards document. It was an attempt to make the existing government standards clearer, and more specific and operational, especially for the teachers. In our case the standards referred to those concerning the content, which in the context of Indonesia meant a more detailed explanation of the standards of performance. The second document listed the characteristics of effective teaching, as identified by EER, especially those based on the classroom factors of the dynamic model of Creemers & Kyriakides (2008).

In the first intervention, one group of teachers used the elaborated standards document, while in the second intervention the elaborated standards document was combined with a teacher improvement program. The teachers in the first intervention group attended a one-day workshop on the elaborated standards document and were then asked to decide themselves upon which effective teaching strategies to adopt. The teachers in the second intervention group also attended this workshop, plus another one-day instruction on effective teaching and six half-day monthly meetings as part of the teacher improvement program. We investigated the effectiveness of these two interventions on the basis of elements provided by EER and performed a multilevel analysis in which we used background characteristics as the controlling variables.

The study was conducted in Indonesia. The reasons for this choice will be explained later. Yet, it is important at this point to briefly illustrate the standards in Indonesia. There are eight, of which two were the most relevant in this study: the standards of graduate competence or performance and the standards of content. The standards of graduate competence describe the expectations concerning the capabilities achieved by the students in the different subjects when they graduate from one level of education to another. The standards of content are a detailed explanation of the knowledge and skills which students are expected to develop in each grade, semester (half-year), and subject domain.

1.2 The Indonesian Context

1.2.1 Demography and general education system

Indonesia is a very large country which consists of about 17,000 smaller islands divided into five big regions. It is the fourth most populated country in the world. According to the 2010 census, it is populated by no less than 237 million people, who are ethnically and linguistically diverse; there are more than 300 ethnic groups and 250 dialects, but one official language, Bahasa Indonesia (Badan Pusat Statistik, 2011).

The Indonesia Education Act (2003) has divided the education system into three age groups: basic education, middle or secondary education, and higher education. The basic education consists of six years of elementary or primary school and three years of junior secondary school, which since 1994 has been declared as a nine-year compulsory education trajectory. The middle or secondary education contains a three-year schooling program at the general or vocational senior secondary school. The higher education prepares for the diploma degree (1–4 years), the bachelor degree, the master degree, and the doctoral degree. Children start elementary school at age 6 or 7, prior to which they go to kindergarten or a play group. The education system is grouped into two categories. The first is general school, which is managed and supervised by the Ministry of National Education (MONE) and the other is madrasah, which is under the authority of the Ministry of Religious Affairs (MORA). Children who attend the madrasah program follow the national curriculum similarly to those who go to general schools, but they have to attend additional religion subjects, which is why they receive more hours of schooling.

Furthermore, based on the Indonesia act of Education, 2003, schools are accredited every five years. Accreditation is an instrument for assessing schools' performance for both improvement and accountability purposes. The maximum score is 100, which is categorized into three scales, where 86 – 100 is labeled A, 71 – 85 B, and 56 – 70 C. Schools whose score is below 56 are not accredited.

1.2.2 Student Outcome

In terms of student achievement, the performance level of the Indonesian students has been low, as indicated by both international and national measurements. The International Mathematics and Science Study – Repeat (TIMSS-R) reported that Indonesia has been performing well below most of its Asian neighbors, as in math it was ranked as the 34th country out of 38 and in science the 32nd (Mohandas, 2004). The results of the Program for International Student Assessment (PISA) convey a similar story. In 2009, Indonesia ranked as the 57th country out of 65 in reading, as the 61st in math and as the 60th in science (Kompasiana, 2011)

At the national level, some less recent studies conducted by Tilaar (1992) and Suryadi (as cited in Sumintono, 2006) indicated that except for civics and Indonesian subjects, students' results were low at that time, on average 6 on a scale from 0 to 10. The current situation, however, is not different. In line with the standards movement, the government launched a centralized national exam to qualify for the requirements of the graduate standards, used since 2003. Thousands of students failed this exam when it was implemented for the first time. In Jakarta alone, the rate of failing students increased to about 16% compared to that of the school exam in the previous year, which was only 3% (Setiogi, 2003). In the following years the trend showed an improvement, but in 2010 the national percentage of failing students in junior secondary school doubled to 10% from 5% in 2009. DKI Jakarta, considered as Indonesia's barometer, was among the five provinces with the highest rates of failing students, which amounted to 29%. In addition, there were 561 schools (1%) where all students (100%) failed (Kompas, 2010).

When the national exam was launched, 3 subjects were tested: Bahasa Indonesia, math, and English. Since the school year 2007/2008, science has been

added. Among the four subjects, English in junior secondary school has produced the lowest results throughout the years (Badan Standarisasi Nasional Pendidikan [BSNP], 2010). This finding is in accordance with a previous analysis by Jazadi (2003), who established that the teaching of English in Indonesian schools has not been generally successful. Furthermore, Jazadi cites several authors, such as Sadtono, O'Reilly, and Handayani, who found that less than 15% of the Indonesian students can be considered to be highly proficient. Another study, conducted by Nurweni and Read (1999), concluded that first-year university students only master 226 English words, which is far below the threshold level of the senior high school completion requirements (4000-5000 vocabulary items). In brief, the student achievement in Indonesia has been low, including in English, and teachers are one of the most responsible parties for this situation.

1.2.3 Teacher Quality and Policies to Improve it

The common classroom in Indonesia is characterized by a low degree of activity. The children sit in rows and are not actively involved in the learning process. Neither are they given enough tasks to keep them busy. They just sit and listen to their teachers. Utomo (2005) explains that the classrooms are characterized by a didactic, whole-class style of teaching. The teachers pay only little attention to the children's needs as individual learners and do not recognize that they have their own ideas, opinions and conceptions about their world. Mostly, a lesson covers only one topic while no links are made with previous lessons, daily life situations or other subject domains. Kaluge, Setiasih, and Tjahjono (2004) argue that the teachers are to some extent to blame for this situation because, as the research of these authors has shown, this group is not capable of creating an active, joyful and effective learning environment.

With respect to the Indonesian teachers' professional qualifications, the current data show that a significant number of primary school teachers has only finished senior high school (417,404 out of 1.25 million teachers) (Kompas, 2008). This information has been confirmed by the statistical data of the Ministry of National Education. It is reported that only 15% of the teachers and principals at the primary schools are qualified of executing the teaching-learning process adequately. At the junior secondary schools the percentage of qualified teachers and principals is

higher compared to that at the primary schools, namely 60% (Departemen Pendidikan Nasional [Depdiknas], 2008). In the case of madrasah the situation is even worse. The Asian Development Bank [ADB] (2006) indicated that more than 90% of the teachers in this segment teach subjects for which they are not qualified in terms of their degrees. These subjects are, for instance, biology, chemistry, English, and math.

Several programs have been developed to improve the quality of these teachers. Some examples are: in-service teacher training [INSET], the Islamic Schools English Language Program [ISELP] in East Java, the Madrasah Education Development Program [MEDP] in East Java, Central Java and South Sulawesi, the JICA Technical Cooperation Project For Development and Mathematic Teaching for Primary and Secondary Education in Indonesia [JICA – IMSTEP] in West Java, Yogyakarta and East Java, the Primary Education Quality Improvement Project (PEQIP) in Aceh, Sulawesi Utara, Sumatera Barat, Yogyakarta, Bali and NTT, and the Continuous Improvement Learning Program in the Jombang district, East Java (ADB, 2006; Cayhono, 2008; Hendayana, 2007; Jazadi, 2003; Van der Werf, et al. 2000).

It is not known whether these programs were implemented with a focus on student achievement as a measurement of their effectiveness and success. The report of PEQIP, however, indicates that after controlling for student prior achievement, it was found that this program had indeed produced some impact on the student achievement level, although small, not always significant, and not in every subject domain (Van der Werf, et al. 2000). In general, however, the development programs aimed at improving teacher professionalism have generally been criticized as ineffective, while the teachers have remained teaching in their usual fashion. Some other problems identified include: large classroom size, heavy teaching loads, insufficient preparation time for the teachers, noisy classrooms due to a lack of soundproofing and an ill classroom design, as well as equipment shortages. Other issues are the pressure of the national exam, which makes teachers take a safe approach to preparing the students for this test, ad hoc programs, the authority of multiple senior staff members, a bureaucratic environment, and the dependence upon external resources (Hendayana, 2007; Nielson, 2003; Thair & Treagust, 2003).

The most recent well-known policy is certification, which is an attempt to measure and control the teachers' quality in terms of their pedagogic, professional/academic, social and personal competencies. Certified teachers have economic benefits, such as additional salary. No studies have been conducted yet into the effectiveness of this policy, but problems are already emerging. The certification of the teachers is based on a portfolio assessment. It is reported, however, that the assessors often possess insufficient knowledge of how to measure teacher quality properly (Winarsih, 2008), which indicates that the guidelines provided by the government are not adequate. Another serious problem is the fact that of the 10 requirements formulated in the portfolio assessment, only two are generally met by the teachers, namely attendance of seminars/workshop and following training/education. Research, presentation skills, and text-book development are activities on which the teachers are hardly focused (Huriyah, 2009). The certification program has currently been put online for accountability purposes. As part of it, the government recently launched a teacher competence test. In 2012, the national average score of teachers teaching in various levels of schooling was approximately 40 – 60 (out of 100) (Suharto, 2012).

Other problems are unequal rates as regards the yearly participation in the certification program between, for instance the state/public and the private sectors; 75-85% of the public organizations, while both groups are nearly equal in size (1,528,472 public and 1,254,849 private) (Huriyah, 2009). The madrasah teachers form the most disadvantaged group, as 90% of the madrasah consists of private school organizations with, of course, private teachers. In short, the teacher quality remains low, but the teachers are certainly not the only ones to blame. However, in spite of the development and implementation of teacher improvement programs, the results have not yet been satisfying. In addition, there are problems associated with the implementation of the certification program, which is likely to have an unfavorable impact on the quality of the teachers as well. Moreover, portfolio as the only way to evaluate teaching quality may well be an insufficient approach to conduct a comprehensive measurement of teaching quality.

1.2.4 Standards-Based Education

The above-described teacher certification forms part of the strategies conducted by the Indonesian government to promote the standards-based education. Under the new law No 20/2003 of the National Education, the government set up eight education standards, two of which are relevant to this thesis: graduate competence/ performance standards and content standards. These two standards have been expected to be useful in steering the teaching and learning processes, including the assessment component.

In line with the standards-based education, the government requires of the schools and teachers to develop their own curricula, syllabi, and lesson plans. These resources have to meet the prescribed standards of performance and content. On the one hand, this measure has increased the autonomy of the schools and teachers. Yet, at the same time it challenges them as regards a number of issues, for example how to develop a curriculum at the school level and how to contextualize the teaching and learning processes in accordance with the immediate environment: there where the students live. Moreover, they used to be merely passive receivers of any curriculum package from the government. Furthermore, the analysis of the standards document carried out in the first stage of this study unveiled some fundamental problems, such as the use of ambiguously broad, inconsistent and general language, and the absence of clear indicators of the competencies that the students are supposed to meet. It is therefore not surprising that many teachers have problems with both taking in and implementing the concept of standards-based education. They simply do not understand it well, while neither the provision of information nor the supply of support resources is adequately taken care of (Hanafie, 2007; Lampung Post, 2006; Prasetiyo, 2009; Suara Merdeka, 2009; Sulistiyani, 2009).

With respect to the curricula development, the general approach among the schools is to “copy and paste” one another’s curriculum documents (Chodijah, 2012, Tuhusetya, 2007). The documents have more of an administrative function than that they serve as a guidance tool for the realization of adequate teaching and learning practices. At the classroom level, as happen in other countries, teaching to the test has become a common practice in Indonesian classrooms (Hendayana, 2007). Finally, in Indonesia there is still little known about the true impact of the standards-based education on the issues of teaching quality and student outcomes.

1.3 Research Problem

The aim of this study has been to investigate how the performance of students in Indonesia can be improved. With respect to the educational problems, the literature briefly reviewed earlier in this chapter has highlighted two improvement approaches: standards-based education and teacher improvement based on the results of EER. Education standards are important in promoting a common understanding, especially among the teachers, of what to expect from the students. However, these standards need to be clear and specific in order to avoid confusion and misinterpretation. As indicated by the findings of EER as regards the important role of the teachers, we know that it is desirable to support this group in improving their quality of teaching, in particular with respect to the factors which have proven to be related to teacher effectiveness. In our opinion, a combination of both approaches might be more fruitful than solely implementing the standard-based education, which until now has been the case in Indonesia. For this reason we studied both the effects of the standards-based approach and those of the combination of this method and a teacher improvement program. We developed two interventions, one in which only the elaborated standards were provided to the teachers, and one in which the elaborated standards were combined with a teacher improvement program. To test the effects of these interventions, we compared them by means of a control group. The research questions are as follows:

1. What are the effects of an intervention solely based on the elaborated standards and those of an intervention in which the elaborated standards are combined with a teacher improvement program on improving teaching quality?
2. What are the effects of an intervention solely based on the elaborated standards and those of an intervention in which the elaborated standards are combined with a teacher improvement program on improving student outcomes?
3. To what degree can the effects of the interventions at enhancing student outcome be explained by the improvement of the teaching quality?

To address the above questions, a one-year longitudinal experimental design involving three different conditions and three measurements was set up. The first condition referred to a situation in which only the elaborated standards were provided to the teachers, while the second referred to a situation in which the teachers were

offered both the elaborated standards and the participation in a teacher improvement program. The third condition was the control situation, where the teachers only utilized the resources made available by the government. To determine the effects of the two conditions, we measured student outcome, both cognitive and non-cognitive, and teaching quality. To examine the cognitive outcomes, the students in the three groups were both pre-tested and post-tested. So we used a pretest-posttest control group design, in which the schools were randomly assigned to the conditions.

In view of the innovation of education, the results of this design have been considered highly important for both policy-making and practice (Slavin, 2010). The pretest has been regarded as relevant for several reasons: 1) it provided a baseline for the status of the groups prior to the intervention, 2) it functioned as a covariate to increase statistical precision, 3) it ruled out threats to internal validity (Smith & Glass, 1987), and 4) it delivered a basis to calculate gain or growth (Jo & Muthen, 2003). Furthermore, combining an experimental with a longitudinal design intensifies the relationship between effectiveness research and educational improvement (Creemers & Reezigt, 1996). A longitudinal design is beneficial because it gathers information about a particular subject at several periodic intervals (Janson, 1990; Magnusson, Bergman, Rudinger, & Torestad, 1991), thereby providing information about issues such as stability and change over time (Gustafsson, 2010; Ruspini, 2002).

1.4 Structure of the Thesis

This thesis consists of six chapters. Chapter 2 provides the theoretical background and chapter three describes the research methodology used in this study. In chapter 4 we discuss our descriptive findings and in chapter 5 we address the principal research questions. Finally, chapter 6 presents the conclusions of our research and suggests some topics for discussion.

In chapter 2, both standard-based education and EER are described in more detail. In the standard-based education section we will go into the history of the concept, its definition, the promises made, the criticism and its effectiveness. The section about EER presents the factors that have been found to affect student outcomes. The final section in this chapter concludes the implications of the standards-based education and EER for the interventions in this study.

Chapter 3 explains the research methodology. It outlines the research design, the research sample, the interventions, the research variables, the research instruments used, and the research procedures. Chapter 4 focuses on the differences among the three sample groups in terms of teaching quality and student outcome. Chapter 5 reports the results of the multilevel analysis in relation to the main research questions. Finally, in the last chapter we present the summary and the conclusions, followed by a discussion, an overview of this study's strengths and limitations, and some comments on the practical and scientific implications of the research.

CHAPTER 2

STANDARDS-BASED EDUCATION AND EDUCATIONAL EFFECTIVENESS RESEARCH;

The Knowledge Bases for Educational Improvement

2.1 Introduction

As far as policy makers are concerned, the ultimate goal of education is to enhance students' learning outcomes, both to improve the low educational quality and to maintain the results attained in this field. To achieve this goal different approaches have been used to increase student outcomes. This chapter presents two theoretical considerations which may form the knowledge base of educational improvement, and which have had implications for the direction of this study.

Firstly, we will review the standards-based education. As previously indicated, research has shown that education standards are used on a worldwide basis to improve education (e.g. Dowson, McInerney, & van Etten, 2007; Neumann, Fischer, & Kauertz, 2010; Schmidt, Houang, & Shakrani, 2009). There has been a strong argument for the notion that education standards enhance student performance through the provision of clear guidelines that help teachers in assisting students to achieve higher performance rates, regardless of their backgrounds. Furthermore, the standards have also been promoted for accountability purposes, based on which schools have to report their results and take responsibility for the possible consequences of their actions (e.g. Baines & Stanley, 2006). In this study, however, the education standards have been considered more as a strategy to improve education than as a tool for increasing accountability, which in itself – for that matter - is also related to educational enhancement. To obtain a more comprehensive understanding of the education standards movement, the first section of this chapter will explain its history, its definition, the promises made, the criticism and its effectiveness.

Secondly, section 2.3 presents the Educational Effectiveness Research (EER). The findings of EER are in line with our argument that education standards are important in providing similar expectations about students' accomplishments, but that it is not enough to work only with standards. Moreover, to date only a few research studies, especially experimental ones, have been conducted to examine the effectiveness of the standards, including in Indonesia. These studies have not provided sufficient unambiguous information on how exactly the problem of the low level of education should be tackled. What remains clear, however, is that educational

improvement is necessary. Moreover, once favorable student outcomes are accomplished, a remaining task for the schools is to maintain these results. EER has identified different factors related to student outcome, thereby providing a knowledge framework based on which improvement measures can be formulated.

The second part of this chapter gives an outline of what works in education. Here, the emphasis is on teacher effectiveness in the light of the findings in this field with respect to the importance of teacher and classroom factors. Furthermore, we will introduce the dynamic model of educational effectiveness (Creemers & Kyriakides, 2008), an approach which has been empirically validated (Antoniou, 2009, Antoniou, Demetriou & Kyriakides, 2006, Kyriakides, Creemers, & Antoniou, 2009) and which is the most recent innovation developed in EER (Reynolds, Sammons, De Fraine, Townsend, & Van Damme, 2011). Finally, the last section of this chapter explains the implications of the standards-based education and EER for the interventions in this study.

2.2. Standards-Based Education

2.2.1 History, Definition, and Promises

The twentieth century has witnessed tremendous educational reforms across the world, one of which has marked the introduction of the education standards. Other examples include the launching of the Soviet satellite *Sputnik* in 1957 (Mathison & Ross, 2008), the establishment of the National Council for Accreditation for Teacher Education (NCATE) in the US in 1954 (Irons, Carlson, Lowery-Moore, & Farrow, 2007), and the publication of the well-known report “A Nation at Risk” in the US in 1966 (Ericson, 2005; Marzano & Kendall, 1996; Mathison & Ross, ibid; Superfine, 2005). The Common Core State Standards Initiative implemented during Obama’s administration, which has replaced the No Child Left Behind act, is the most current standards movement in the US (Schmidt, et al. 2009). Although the literature on the standards movement is dominated by US publications, the movement has also spread to various countries in Europe, Asia and to Australia (e.g. Choi, de Vries, & Kim, 2009; Delandshere & Petrosky 2003; Faizi, Shakil, & Lodhi, 2011; Neumann, Fischer, & Kauertz, 2010; Schmidt, et al. 2009; Widmer, 2004). In brief, the standards movement has currently been one of the strongest leading forces and is labeled

"the hottest item" in the global educational reform (Chalker & Haynes, 1994; Cizek, 2001; Lewis, 1995).

Various problems have been identified which justify the development of education standards. They include (1) low student achievement rates (e.g. Education Commission of the States, 2002; National Commission on Excellence in Education, 1983; Neumann, et al. 2010; US Department of Education, 2008; Tienken, 2011), (2) the continuing differences in student achievement in terms of background, such as race and ethnicity (Apthorp, et al. 2001; Ravitch, 1995), (3) the failure of the 1960s movement, which was far too ambitious, containing multiple instructional objectives hard to process by the teachers (Popham, 1997), (4) differences in curricula, assessment methods, and grading systems (Chambers & Dean, 2000; Marzano & Kendall, 1996b; Ravitch, ibid), and (5) a lack of focus on the educational output, whereby the educational input, in terms of time, energy, and money spent, was found to contribute little to improving educational quality and equality (Coleman, et al. 1972 as cited in Marzano & Kendall, ibid).

Another element which has played a role in the standards movement is the global economic competition (Burke & Marshall, 2010; Gibbs & Howley, 2000; Mansilla & Riejos, 2007; Markowitsch & Luomi-Messerer, 2007; Ravitch, 1995). In the US, education standards are considered as an important instrument to ensure financial security and economic competitiveness (Marzano & Kendall, 1996b). In addition, Markowitsch and Luomi-Messerer (ibid) have argued that in Europe there is a close connection between the establishment of the European Qualification Framework (EQF) and the unification of a number of countries into the European Union (EU), whose main objectives have been to make Europe politically more stable, economically more competitive, and socially more coherent. Thus, the call for education standards is not only the result of national but also of international pressures, while their purpose not only has an academic but also an economic and political footing.

In practical terms, education standards are defined as guidelines with respect to students' learning objectives (Dowson et al. 2007; National Research Council [NCR], 2001; Organization for Economic Cooperation and Development [OECD], 1995). Based on a more wide-ranging perspective, Echevarria, Short and Powers,

(2006), Spillane (2004), and Wheelock (1995) have defined these standards as a comprehensive framework for the various activities in the field of education, including all facets from the development of the learning goals, the curriculum design and the assessment methods to the development of teacher quality programs. In line with these views, many types of standards have been defined and various terms are being used. The types include performance standards, content standards, teacher standards and teaching process standards. The terms are, for example, "standards-based education" (e.g. Chambers & Dean, 2000; Goodwin, 2003; Marzano & Kendall, 1996a), "standards movement" (e.g. Dowson, et al. 2007) and "standards-based reform" (e.g. Education Commission of the States, 2002; Hamilton, Stecher, & Yuan, 2008). According to the literature in this field these three terms mean the same, and are therefore used interchangeably in this thesis.

The standards of performance and content are the most relevant ones in this study because they describe the expectations with respect to students' performance. Therefore, both are dealt with. The first one specifies the levels of proficiency that students are expected to achieve (Popham, 1997; National Council on Education Standards and Testing [NCEST], as cited in Hamilton, et al. 2008; Wanacott, 2000). The second one defines the particular knowledge, skills, and insights that have to be taught to make the students attain the proficiency levels specified (Elliot & Thurlow, 1997, NCEST, ibid).

Concerning the promises, by providing a clear direction to what students are expected to learn and accomplish, it is argued that the standards provide a vision on teaching and learning which can serve as a guiding principle to be used by the various education stakeholders, such as teachers, schools (principals), and education administrators at the district/regional levels. This vision has to convey that all students, regardless of their backgrounds, should be provided the same learning opportunities to meet the targets formulated in the standards. In other words, the standards concretely define the goals and directions to be aimed for by the stakeholders (Fraser, 1996). In this respect, the goals and strategies agreed upon have proven to enhance the educational organizations' capabilities in rational planning and action (Susan Rosenhotlz, as cited in Schomoker & Marzano, 1999).

In this way both excellence and equity are claimed to be simultaneously addressed in the field of education (Sandoltz, Ogawa, & Scribner, 2004; Buttram & Waters, 1997; McClure, 2005; Wurtz, et al. 1993). Excellence is expected to be achieved by raising the intellectual rigor through the development of more intellectually demanding content and the development of students' thinking skills (MsClure, ibid; Sandoltz, et al. ibid; Wheelock, 1995). Apart from making the educational content more challenging, standards are meant to clarify which knowledge and skills are meaningful and relevant to learn (Marzano & Kendall, 1996a). Formerly, teachers of a particular subject may have differed in what they taught the students and what they expected of them. They may have based their teaching either on textbooks or/ and on their own teaching materials, resulting in different content and accents. Standards make it possible to discuss all sorts of issues, such as the transparency of the goals, the instruction guidelines, and the efforts required by teachers to help poor and disadvantaged student groups (Sandoltz et al. ibid; Buttram & Waters, ibid; McClure, ibid; Wurtz, et al. ibid).

As regards equity, it is important in this context to have teachers set the same requirements for all students in achieving the targets formulated by the standards. Poor and disadvantaged students should not be used to justify low expectations. Until now, teachers may have had different expectations for different students, classes, and schools. Using standards, teachers are expected to guide all of their students in meeting the same minimum requirements; "all students can learn" (Chambers & Dean, 2000).

Standards require the teachers to recognize the students' diverse background characteristics in relation to their educational needs, which could be considered as the greatest challenge (Baines & Stanley, 2006). In order to address these needs, they will have to deepen their knowledge and skills related to their subject domain as well as adjust their instructional strategies (McClure, 2005; American Federation of Teachers, 2009). Concerning instruction, the standards movements advocates a constructivist approach as the basis for the teachers' instructional adjustments. Teachers are expected to abandon their behaviorist approach to teaching, and opt for a method focused on creating a situation in which students are no longer the passive receivers of information but active participants in the learning process

(Corcoran, 1995). Teaching may, for instance, emphasizes cooperation among the students in groups rather than the use of the didactic model of instruction (NCTM, 1989, 1991, 2000, as cited in Snow-Renner, 2001). Thus, the approach conducted should be participatory and student-centered rather than teacher-directed, which is promoted by a clear emphasis on classroom activities (Thompson, 2009).

With respect to the schools, the standards can function as a mechanism whereby these organizations can be held accountable for the contents of their curricula to the relevant stakeholders, such as the students themselves, the parents, the teachers, the education administrators, the policy makers and the public (Raizen, 1998). In this way, successful schools will be rewarded by getting recognition and being considered as model institutions, while less successful (or failing) schools will be stimulated to improve, or are offered help (Lefkowits & Arens, 2004). This form of public reporting has been considered as "an advance" since in most other accountability systems insufficient attention is paid to these stakeholders (Goodwin, 2003).

2.2.2 The Effectiveness of the Standards

Until now, not many studies have focused on the impact of educational standards on student outcomes, especially not those based on experimental designs. The existing studies are largely from the US, and particularly deal with math and science (US Department of Education, 2008). Generally, the findings of these studies suggest that standards-based instruction (SBI) only influences the student achievement rates of a small number of certain groups (e.g. Lauer, Snow, Martin-Glenn, Van Buhler, Stoutemyer, & Snow-Renner, 2005; NRC, 2001; Thompson, 2009). Thompson (ibid) for instance, reports that SBI practices contribute significantly to female student math achievements but not to those of male students, and that regardless of gender, the effect is larger for White students than for ethnic minority students (4% and 2.5% respectively). However, the findings also suggest conflicting results. Lauer et al. (ibid) reviewed a number of studies and found that the student scores of low-performing schools significantly increased after the introduction of SBI, whereas those of high-performing schools rose only very slightly. However, other studies in the review indicated that high-performing students gained more from this approach than low-performing students.

In Germany, the establishment of the National Education Standards (NES) as a response to the significantly poor student performance rates in science, reading and math in PISA 2000, contributed to a general trend toward improvement from 2003 to 2006 (Neumann, et. al, 2010). Furthermore, Neumann et al. (ibid) report that their student performance rates outnumbered the average score of the OECD in science. However, this result was realized by an increase in the achievement of students who had entered the upper proficiency level, whereas the achievement of students on the lowest level of proficiency had remained the same. This finding is in line with the above review which suggests that high-performing students gain more from performance measures than low-performing students.

It seems therefore as if standards only work for certain groups. Low performing students who usually come from minority and disadvantaged groups have not benefited from the movement. Standards-based education has even been considered a reason for members of minority groups to leave the advantaged institutions and move to the disadvantaged schools, which again has resulted in a growing inequity of educational opportunities (Zuzovsky & Libman, 2006). Similarly, the US Department of Education (2008) has admitted that the performance of American students at the high school level is still the same, if not worse, compared to when 'A Nation at Risk' was published, which marked the need for standards movement, and that the situation is much worse for students from minority backgrounds.

Additional research studies into the standards-based education have also identified different types of problems, namely philosophical, physical and practical. These problems may largely explain the ineffectiveness of the standards. Philosophical criticism has pointed at the fundamental consequences of the "one-size-fits-all" approach of the standards-based education (Baines & Stanley, 2006). Problems in a physical sense refer to the language of the standards document, and the practical issues concern the preparation and the implementation stages. Because the philosophical issue exceeds the scope of this study, we have only addressed the last two types of problems, arguing that some of this philosophical concern may already be solved by improving the preparation and implementation stages.

With respect to the physical problems, first, the standards documents are written in broad and general terms (e.g. Choi, et al. 2009; Dowson, et al. 2007; Hammer,

1998). In addition, they are vague and ambiguous, while their content lacks a sufficient level of grounding (Gandall, 1996; Finn, Petrili, & Vanourek, 1998). Furthermore, concerning the instructional strategies, there are no illustrations or examples in the documents to educate the teachers on how to interpret the standards in the classroom; the standards merely consist of lists of topics and do not explicitly guide the teachers in interpreting and applying the instructional strategies (Hill, 2001).

Secondly, from a practical point of view, the plans are often not realistic and ambitious, while there is hardly any focus on teacher involvement in the formulation and planning stages of the strategies (Hammer, 1998). Furthermore, Chamber and Dean (2000) found that the importance of discussing the content of the documents is not always recognized by the people or institutions in charge. Instead, teachers receive the copies of the documents and are asked whether in their view the content corresponds with what they have been doing. Rather than being encouraged to participate in a critical review of the standards, they are merely given the task of discussing them with their colleagues to formulate the subject's curriculum across the different grades (Mathison & Freeman, 2008). Finally, the standards-based education is costly (Baines & Stanley, 2006).

Consequently, problems emerge. Although it has been recognized that standards could be useful in realizing a common educational direction that serves as a guideline for all parties involved (the teachers themselves, the schools, and the district administrators), the teachers have found it difficult to adapt to the standards policies (Chamber & Dean, 2000; Wise & Darling-Hammons, 1983). Moreover, the standards' content is based on the "anything goes catch-all" principle (p. 43), which means that the materials and instructions are too much open to interpretation (Li, 2007). The teachers have argued that they should be more involved in the planning and implementation stages. It is clear, however, that the teachers agree that serious changes need to be introduced in the educational system (Hammer, 1998).

Mathison and Freeman (2008) report that a number of professional organizations in the US, such as the National Council of Teachers of Math and the National Council of Teachers of English, developed some curricular guidelines for what to teach and when. However, their impact has only been moderate because these organizations

did not have authoritative power. Furthermore, as the authors (ibid) explain, many schools (usually those with strong leadership in middle- and upper-class communities) are already engaged in a curricular alignment project to map out how their former teaching activities match with what is prescribed by the standards. In addition, they use special textbooks to guide the teaching-learning processes.

So what is it that teachers do in their classrooms? They have been reported as spending less time on helping the students understand the topics profoundly and focusing more on making sure that the extensive spectrum of topics is covered (NRC, 1996). This approach may be the result of the pressure of the nation-wide assessment policy. Other research has indicated that both the instruction rules and the curricula have been narrowed down to the 'teaching to the test' principle. As a result, learning has become primarily centered around test preparation (Doherty, 2001; Hunt, Jr., Rizzo, & White, 2009; Proefriedt, 2008; Richman, 2001; Wise & Darling-Hammond, 1983). This has also been the case in the Indonesian classrooms (Hendayana, 2007). Proefriedt (ibid) argues that there is a tendency to "copy" the questions presented in the tests. In a survey conducted in 2002, Doherty (ibid) found that 70% of the teachers admitted that their teaching was highly influenced by the state exam and that they spent a great deal of time on teaching their students test-taking strategies. In a similar vein, Baines and Stanley (2006) claim that also principals admitted that the state exam has become the criterion for the content, delivery and timing of the curricula, which surprisingly was considered by the test makers and reformers to be a good development.

To continue, neither have studies on the impact of the use of standards-based curricula (SBC) on teacher instruction yielded good results. In their review, Lauer et al. (2005) found that SBC has been expected to motivate and help teachers in changing their instruction by focusing more on problem-solving activities and pair work while spending less time on presentation and whole-group work. However, the review highlights the continuing use of the more traditional strategies, such as seatwork. Furthermore, the teachers complain about the time-consuming tasks of the preparation and delivery of the teaching materials. Thompson (2009) found that the classroom practices are dominated by the use of strategies such as whole-class lectures, independent seatwork, quizzes and written homework. Thus, it can

be concluded that the teaching approaches have remained old fashion. Standard-based practices, such as inquiry, problem solving, and cooperative learning are rarely being practiced by the teachers.

Another interesting result from the review of Lauer, et al. (2005) is, however, that teachers' professional development was found to be significantly and positively related to teacher instruction and student achievement. Another element addressed by the review was time, which supports the finding of Florian (1999). Teachers need continuous support, and the realization of changes in teacher instruction take time. The review concludes that implementation of SBC alone did not influence changes in teacher instruction. An issue which justifies the concerns regarding teachers' professional development is that the standards movement has been focused on summative results, which are mainly used for accountability purposes and not for providing feedback to the teachers to improve their methods of instruction (Snow-Renner, 2001). This issue indicates the need for building a mechanism which provides the teachers with the necessary assistance and feedback to improve their instruction techniques. Therefore, investing in their professional development is a significant step toward equipping teachers with the necessary knowledge and skills to meet the learning needs of all students in an adequate manner (e.g. Rowe, 2007).

In summary, the findings with respect to the effects of educational standards on both student learning outcomes and teaching quality have proven to be inconsistent and therefore ineffective. At least two factors could be argued to contribute to these results. The first is the fact that the standards documents are often written in a general, vague and ambiguous language. The second is that too little attention is being paid to the teachers, who can be considered as the most important actors in any educational improvement measure. The above discussion has shown that teachers face a great deal of problems. They play, however, the principal role in this area of study, because they form the primary source of student learning in the classroom.

At this point, it is therefore relevant and useful to look at the Educational Effectiveness Research (EER), which deals with factors related to student outcomes. The findings of this research offer both theory-driven and empirical-based prioritization tools for attaining educational improvement. A review of EER,

which is considered as the second theoretical consideration in this study, will be presented in the following section.

2.3 Educational Effectiveness Research

2.3.1 What works in education?

EER aims to identify factors in education that are related to student outcomes and that could be improved by means of intervention programs. The publications of Coleman et al. (1966) and Jencks et al. (1972) on the equity in student achievement and educational opportunity have commonly been considered as marking the beginning of EER (e.g. Creemers, 1994; Doolaard, 1999). Based on their their findings, these authors have attributed a major role to student background characteristics, which led to conclude that schools do in fact not make a difference. Therefore, the next agenda in this field was to open the "black box" of education, which resulted in lists of effectiveness enhancing factors.

These factors, as has been argued, can occur at all different levels, such as school, teacher/classroom, student, context and national policies. In addition, they are considered to influence the outcomes of education both directly and indirectly, while they could be either changeable or unchangable. Identifying these factors, especially the changeable ones, could enable policy makers to define actions of improvement. In this way, a link between research/ theories and improvement practices could be estabhished. The effectiveness research could provide theories on school improvement while the schools could serve as a platform for researchers to test their theories (Creemers, Stoll, Reezigt, and the ESI Team, n.d).

Student outcome has been wildely used as an indicator of effectiveness. Creemers and Scheerens (n.d) and Scheerens and Bosker (1997) define effectiveness as the extent to which planned goals are achieved. Doolaard (1999) argues that the attainment of goals should be controlled for relevant background characteristics, and that effectiveness is more than just a snapshot made at a particular moment in time. In education, student performance, especially its cognitive component, has been widely used to measure the effectiveness of various factors. However, using the cognitive domain as the only indicator of effectiveness has been criticized to narrow down the scope and the meaning of education (Creemers &

Kyriakides, 2008; Van der Werf, Opdenakker, & Kuyper, 2008). However, Haanstra (as cited in Creemers and Kyriakides, 2008) has shown that the impact of education in non-cognitive domains is quite small. Therefore, even though it may be considered as a traditional approach, the use of students' cognitive performance as an indicator of effectiveness is still considered as significant.

Nevertherless, the attainment of non-cognitive goals, such as student motivation, should continue to be promoted. In the field of educational effectiveness research, motivation has been regarded as a non-cognitive outcome as well as a predictor of cognitive outcomes, among other predictors such as intelligence and prior achievement (Van der Werf, et al. 2008). However, because motivation has appeared to be an unstable factor, and the research into motivation as a predictor of academic achievement has yielded unclear results (Van der Werf, et al. ibid), it is arguable to consider motivation as a non-cognitive outcome. There are certain factors that cause motivational variance and instability, which is supported by the educational psychology theory. Thus, educational effectiveness research concerns the attainment of student learning outcomes in both the cognitive and the non-cognitive domain, after controlling for relevant background characteristics.

What has the research been saying in the past decades about factors closely related to student learning outcomes? Several studies have listed a number of factors which could be classified into different categories, such as school, teacher, student and context/country policy (e.g. Creemers, 1994; Creemers & Kyriakides, 2008; Muijs & Reynolds, 2011; Scheerens & Bosker, 1997), as well as changable and unchangable factors. Doolaard (1999) reviews three waves of educational effectivess research which differ in terms of their research designs, statistical models, operationalization of school factors, education systems and sectors, and school effect measures used. However, she rocognizes that there is consensus on a number of factors. Given the concerns of EER with respect to finding factors that could be improved by intervention programs as described at the beginning of this section, our focus will be more on changeable factors.

A number of authors, either based on their own studies or on their reviews of other studies, have listed some common factors which mostly occur at both the school and the classroom levels (e.g. Creemers, 1994; Muijs & Reynolds, 2011;

Purkey & Smith, 1983; Scheerens & Bosker, 1997). Some examples are high expectations, curriculum quality/opportunity to learn, structured versus adaptive instruction, school climate, classroom climate, classroom management, feedback and reinforcement, parental involvement and purposeful leadership.

With respect to the first three factors, Scheerens and Bosker (1997) reviewed studies conducted by Brookover and Lezotte (1979), Weber (1971), Glenn (1981), and Rutter, et al. (1979), concluding that fostering the high expectations as regards student achievement by setting up challenging standards have been found to be one of the most important effectiveness enhancing factors. The mastery of basic subjects as well as records of student achievement have been included in this variable. Furthermore, the curriculum has to contain relevant information on what students will have to learn, thereby providing an outline of students' opportunities for learning. More structured curricula containing an explicit list of goals presented in a hierarchical manner are considered to be more effective than less well-defined educational programs (Creemers, 1994), the effect size of this variable being moderate (0.30) (Kulik and Kulik, 1989). Finally, the most important issue of concern has been the maximization of learning time (Sammons, Hillman, & Mortimore, 1995). These three variables can play a role at both the school and the classroom levels, which indicates that improvement efforts should be focused on these variables at these two levels.

Concerning the school climate, some regularly examined variables include 1) an orderly atmosphere where the rules and regulations are clearly laid down, as well as the policies regarding rewards and punishment, absenteeism and drop-out, and 2) good internal relations among the different parties within the schools, such as the students, the teachers, other staff members, and the head teachers (Scheerens & Bosker, 1997). Scheerens and Creemers (1989) found that school climate is one of the effectiveness enhancing factors. The next is classroom climate, referred to by Creemers and Kyriakides (2008) as the building of the classroom into a learning environment. This factor deals with the role of the teacher in issues of classroom management, such as the creation of a safe and order learning environment in which student's opportunities to learn is maximized. It has been commonly found to be positively related to student outcomes (e.g. Creemers & Reezigt, 1996; Muijs & Reynolds, 2000).

Another factor strongly related to student outcomes is leadership, particularly instructional as opposed to administrative, as observed by e.g. Scheerens (1992) and Teddlie & Stringfield (1993). However, due to inconsistent results, which could be explained by differences in school settings and the roles of principals as well as the use of different research theories and methodologies, the relevance of this factor has been questioned (Doolaard, 1999). Furthermore, also more physical-related conditions, such as school location and size have been investigated. Some studies (e.g. Kaluge, Setiasih, & Tjahjono, 2005; Van der Werf et al. 2000) found that students in urban areas achieved better than those in rural areas. With respect to school size, which in some cases is referred to as class size, the current research findings are rather complex. Although there is evidence that small classes are beneficial for reading and math, there are also signals that low achieving students tend to improve more in smaller classes (Block, 1985). Plecki (1991) found that larger schools are not associated with better student achievement. In addition, this study established a negative relationship between small school size and student achievement in the case of low SES pupils. Furthermore, student-teacher ratio has been found to be the most consistent school-level predictor, whereby a higher student-teacher ratio is negatively associated with reading, writing, and math (Hoyle, O'Dwyesr, & Chang, 2011).

Concerning student motivation, the effect of the school has unexpectedly been found to be minimal. Van Damme, Opdenakker, Van Landeghem, De Fraine, Pustjens, & Van de Gaer (2006) reviewed a number of studies and concluded that schools have a larger effect on student achievement than on non-cognitive outcomes such as student motivation. This finding suggests that other factors, which might occur at different levels, especially the individual, play a more important role. In addition, the authors note that the research into the effects of schools on non-cognitive outcomes is scarce.

Apart from the above-mentioned school and classroom factors, another item which largely contributes to student outcomes has appeared to be student background characteristics. Coleman et al. (1966) and Jencks et al. (1972) have shown that initial differences in ability, SES, and gender play a major role in the variance of the learning outcomes. Also more recent studies have reported on this

trend, presenting figures from 75% up to 83% (Bosker as cited in Creemers, 1994; Stringfield & Teddlie, 1989). With respect to gender, studies conducted in different countries in America, Asia, and Europe have consistently indicated that girls achieve better in reading (Education, Audiovisual and Culture Executive Agency, 2010; Fleischman, Hopstock, Pelczar, & Shelley, 2010; Shafiq, 2011; Van Damme et al. 2006). As regards student motivation, it has been found that boys' motivation is generally lower than that of girls for learning to read (Baker & Wigfield, 1999; Wigfield & Guthrie, 1997), and that the role of SES is rather mixed, with a small correlation found in the lower grades while no relationship was observed in the higher grades (Hustinx, Kuyper, Van Der Werf, & Dijkstra, 2009). Wong (2007) argues that peers form the most influential factor in student motivation.

However, although the role of student level factors in learning outcomes is considerable, it is also important to gain an insight into the influence of other domains. Only then can the improvement efforts be properly prioritized. Moreover, also the higher levels, i.e. schools and context or region, play a role in defining effective practices. Some authors (e.g. Bosker & Scheerens, 1994; Creemers, 1994; Creemers & Kyriakides, 2008) have acknowledged the role of context in the realization of effective practices on both the school and the classroom levels. Effectiveness is therefore multi-leveled and multi-faceted in nature. In addition, student factors such as SES and gender are unchangeable, which means that they cannot be influenced. Creemers (1994) concludes that after controling for student background characteristics, school and classroom characteristics explain about 12-18% of the variance in student outcomes. However, when separated, the effect size of the classroom characteristics is clearly larger (Luyten & Snijders, 1996).

The superiority of the classroom or teacher factors has been widely recognized across countries (e.g. Creemers, 1994, Darling-Hammond, 1997; Harris &Muijs, 2005, Hill & Rowe, 1996; Luyten & Snijders, 1996; Marzano, 2007; Van der Werf, et al. 2000), although it is dominated by certain subjects, such as math and language (Doolaard, 1999; Muijs & Reynolds, 2000). As argued by Doolaard (ibid), this superiority could be the result of the fact that classroom teachings as well as the learning process are the primary sources of learning. Although the effect size of classroom teaching is relatively small compared to that of student background

characteristics, it plays an important role in the curricular and development planning schemes, especially in view of the objective to influence all students (Luyten, 1994). The question is then how extensive the effect of classroom or teacher factors on student learning outcomes really is, which will dealt with in the following section.

2.3.2 Teacher Effectiveness

To start the discussion, it is important to first consider the definition of quality or effective teaching. There has been little consensus on the definition of this concept (Darling-Hammond, 2000; Seidel & Shavelson, 2007). It has even been claimed that it is impossible to define teacher effectiveness, because the criteria used among the teachers and other parties involved (Perrott, 1982) for the subjects, grades, the locations and the available resources (Needels & Gage, 1991) vary to such a large extent. In addition, the range of items or variables investigated is extensive, such as teacher beliefs, subject knowledge, behavior, and teaching artistry (Harris, 1998; Muijs, 2006). Moreover, teaching quality has been measured from many different perspectives, such as teacher and student perception (e.g. Money, 1992), classroom observation (e.g. Hill, et al. 1991), and the "process-product" paradigm, referring to the extent to which teachers' behavior in the classrooms are related to student achievement (e.g. Antoniou, 2009; Borich, 1996; Brophy, 1981; Lavy, 2011; Kane, Taylor, Tyler, & Wooten, 2010). All of these approaches have resulted in a considerable number of different definitions.

A survey of student and teacher perceptions of effective teaching has yielded the following list of characteristics of effective teachers: 1) knowledge of the subject matter, 2) effective communication, 3) well organized materials, 4) ability to motivate and inspire, 5) being friendly and open, and 6) classroom control (Money, 1992). Another finding concerned the confirmation of a small positive correlation between teacher professional qualification and student achievement, which is closely related to the knowledge of the subject matter (Borich, 1992; Darling-Hammond, 2000; Heck, 2007; Monk, 1994). Monk (ibid) adds, however, that the relationship is curvilinear: teachers require a minimal level of knowledge to be effective, but beyond a certain point, a negative correlation occurs. Teaching experience is another factor which has been extensively investigated (Scriven, 1994; Reynolds & Muijs, 1999).

Creemers (1994) argues, however, that teacher experience cannot be influenced. Therefore, this variable cannot be adapted with the aim of improving teacher effectiveness.

Muijs and Reynolds (2011) present a review of the empirical evidence gathered in studies conducted during the 1970s, 1980s, and 1990s. They found that almost 60 different teacher behaviors are associated with student outcomes, which indicates that effective teaching is basically considered to be determined by a large number of small items rather than by a small number of large components. This process-product paradigm has been regarded to be a better approach than other methods in defining the concept of effective teaching. It provides information on teacher behaviors and on how these behaviors differentiate among student outcomes. In brief, this method offers empirical evidence, which is a strong and important basis for defining and prioritizing teacher improvement efforts. Based on the above, therefore, in this thesis the process-product paradigm was chosen to define the concept of teaching quality.

Compared to other factors or variables, the actions of teachers in the classroom have been found to explain a large proportion of the classroom level variance (Creemers, 1994; Creemers & Kyriakides, 2008; Muijs & Reynolds, 2010). More specifically, teacher instruction – for which different terms are being used - has been largely recognized as the most important aspect in teacher behavior, both in terms of quantity and quality (e.g. Brophy & Good, 1986; Marzano, 2000; Powell, 1980; Wang & Walberg, 1991; Muijs & Reynolds, 2011). Here teacher instruction refers to teachers' efforts in maximizing the students' opportunities to learn (Creemers, 1994; Rosenshine, 1983). Several activities have been identified and empirically validated to promote students' learning opportunities, such as an emphasis on academic goals and achievement (Cotton, 1995; Doyle, 1986; Powell, ibid), a clear and step-wise presentation of the materials, effective questioning and feedback (Bennett, Desforgess, Cockburn, & Winkelson, 1981; Brophy & Good, ibid; Kane, et al. 2010; Muijs & Reynolds, 2000; Scheerens & Bosker, 1997; Seidel & Shavelson, 2007; Smith & Land, 1981; Wragg 1984; Doyle, ibid), and clear structures and routines (Soar & Soar, 1979).

Concerning the emphasis on the academic objectives, several studies have shown that by clearly and explicitly explaining these goals, learning becomes relevant and meaningful, which increases students' motivation (e.g. Althoff, Linde, Mason, Nagel, & O'Reilly, 2007; De Corte, 2000; Paris & Paris, 2001; Padak, 2002). Based on several studies (e.g. Kulik & Kulik, 1989; Mager, 1962), also Creemers (1994) concludes that explicit goals, particularly when listed in a hierarchical order, contribute to effectiveness. Furthermore, Stallings, Corry, Fairweather, & Needels (1978) suggested that effective teachers should particularly focus on the academic activities involved in achieving the goals rather than on having students work on their own. Effective teachers also provide sufficient practice, while they monitor the students' work and provide appropriate feedback (Bohn, Roerig, & Pressley, 2004). Creemers and Kyriakides (2008) have labeled this practice as application activities, aimed at offering students the opportunity to instantly exercise the lesson material and at providing both individual students and groups with the proper feedback. Concerning feedback, it was found that effective teachers encourage especially low-SES and low-achieving students more frequently in terms of student effort (Kyriakides & Creemers, 2006).

Next, effective teachers present the subject materials in relation to the goals in a stepwise manner, organized into small parts (Joyce, Weil, & Calhoun, 2000). Some studies reviewed by Brophy and Good (1986) have claimed that students achieve more when teachers structure the materials by 1) offering an overview or review of the objectives, 2) outlining the contents to be covered and signaling the transitions between the lesson parts, and 3) calling attention to the main topics at the end of the lesson. When starting the lesson, effective teachers review or practice the items learnt in the previous lesson, for instance by going over the homework (Muijs & Reynolds, 2011; Rosenshine & Stevens, 1986). In this way they make a link between what students have learned and what they are about to learn in the lesson to come. Simons (as cited in Creemers, 1994) noted that the results of including homework in the curriculum are encouraging, especially for disadvantaged groups. Similarly, Van der Werf (1995) also found that homework had a positive effect on student achievement.

When outlining (or structuring) the contents of the material, effective teachers consider the length of time needed for each item, while recognizing the importance of an appropriate pace. Especially for primary students and in the teaching of basic skills, a higher tempo is recommended (Smith, Hardman, Wall, Mroz, 2004) as it retains the momentum and interests of the students and allows more contents to be covered (Muijs and Reynolds, 2011). In addition, elements of structuring are argued to connect the different items and activities of the lesson (Case, 1993). Then, reviewing or repeating the main topics will lead to a degree of redundancy in the information, which has been found to increase student achievement (Leinhardt, Weidman, Hammond, 1987; Smith & Sanders, 1981).

Furthermore, effective teachers guide classroom discussions through questioning (Muijs & Reynolds, 2000). In reading for instance, questioning has generated higher achievement rates (Kane, et al. 2010). The questions vary in terms of difficulty level and types (process and product) in accordance with the objectives, where product questions require specific answers and process questions the use of processes or procedures (Muijs & Reynolds, 2011). In terms of difficulty level, some studies have recommended teachers to create exercises of which 75% of the questions are expected to be answered correctly by the students (Anderson, et al. Brophy & Evertson, as cited in Creemers & Kyriakides, 2008). Similarly, Brophy and Good (1986) suggest that teachers should include 25% of high level questions in each exercise. With respect to the difficulty level of questions, effective teachers take the different contexts into account. Teaching basic skills, for instance, requires both a great deal of drill and practice accompanied by frequent fast-paced review, and a rapid delivery of the correct answers. However, when teaching complex cognitive skills such as generalization and evaluation, often only a few students may answer correctly. And sometimes there may be several correct answers (Brophy & Good, 1986).

Considering the types of questions, effective teachers particularly focus on process questions (Askew & William, 1995; Brophy & Good, 1986; Everston, Anderson, Anderson, & Brophy, 1980; Muijs, Chapman, Armstrong, & Collins, 2010). Furthermore, effective teachers equally distribute the questions and provide appropriate feedback to the answers, for example by providing hints or clues or

formulating the questions in an easier manner if the students cannot answer them. Finally, the questions posed by effective teachers are always clear, and they give the students sufficient time to answer them (Creemers & Kyriakides, 2008).

In addition to the instructional aspects described above, other necessary *ingredients* for the enhancement of student outcomes are classroom management and high student engagement. This finding has been confirmed by a number of studies, such as Brophy & Good (1986), Doyle (1986), Creemers & Reezigt (1996), Muijs & Reynolds (2000, 2011) and Powell (1980). Classroom management, sometimes referred to as the classroom climate, is associated with the behavior of the stakeholders (Heck & Marcoulides, 1996), which according to Doyle (1986) involves two aspects, namely learning and order. Learning is the instructional component and order is the managerial component. Creemers (1994), however, points out that although management is necessary to facilitate learning, it is in itself not sufficient to improve student outcomes. Effective teachers spend more time on academic activities than on managerial tasks (Powell, 1980).

Additionally, it is useful to look at the curriculum and other planning documents, as they give an impression of how the teacher planned his/her lessons to provide the students with the optimum opportunities for learning. Although Van der Werf (as cited in Creemers, 1994) found that the influence of planning documents, such as the curriculum, the school working plan, and other activity plans, is much smaller than the other factors, it is logical to assume that teachers derive their teaching-learning processes from the goals they plan. It has therefore been suggested that effective teachers plan their lesson well. However, basing one's learning goals only on the planning document is not sufficient, because teaching is not a straightforward process (Creemers, ibid). It is often the case that because of several reasons a teacher cannot strictly adhere to his/her plan. Hence, to fully understand their educational goals, it is important to understand teachers' educational choices and activities in the classroom.

Finally, in order to contribute to the development of the teacher effectiveness theories, Creemers (1994) developed a model of the effective classroom, which sums up the different factors of instruction into three factors, namely curriculum, grouping procedure, and teacher behavior. Several studies have examined the validity

of this model (e.g. De Jong, Westerhof, & Kruiter, 2004; Kyriakides, 2005; Kyriakides, Campbell, & Gagatsis, 2000). It is considered as one of the most influential theoretical constructs in the field (Teddlie & Reynolds, 2000). Moreover, the model has been further developed into a dynamic model of educational effectiveness (Creemers & Kyriakides, 2008).

This dynamic model has four levels: context/national policy, school, teacher/classroom, and student. Similar to other effectiveness models, its structure is therefore multileveled. The teacher/classroom level is emphasized in this model, while the context and school levels are expected to provide the conditions necessary for maximizing its effectiveness. Responding to the criticism that effectiveness models usually fail to provide the means to measure all effectiveness factors adequately, the dynamic model has proposed five dimensions to measure each one on each level separately: frequency, focus, stage, quality and differentiation.

Frequency refers to the quantity of the activities associated with the effectiveness factors. Focus concerns the specificity of the activity: whether it is too specific or too general in relation to its goal. Stage relates to the particular moment in time when an activity takes place. In this context, there may be activities which do not necessarily increase student outcomes. It is therefore important to regularly assess their efficacy, for instance by critically considering the tasks defined and/or checking whether they are supported by the literature. Another relevant aspect is whether the students can understand and perform the activities. Finally, differentiation concerns the diversity of the subjects involved, which in the case of the classroom are the students. Classrooms usually include different groups of students. Teachers are expected to address these groups in such a way that all are provided with equal opportunities in the classroom.

Based on the findings of the above-mentioned teacher effectiveness studies, the model has linked the observable teacher instructional roles to student learning outcomes. Eight factors have been defined at the classroom level: 1) orientation, 2) structuring, 3) questioning, 4) teaching modeling, 5) application, 6) management of time, 7) the classroom as a learning environment (CLE), and 8) classroom assessment. ‘Orientation’ concerns the explanation of the objectives, which is expected to help the students understand the importance of their learning activities.

'Structuring' refers to the explanation of the distribution of the series of activities of the lesson. 'Questioning' entails the attempt to categorize the questions in terms of difficulty level and type (product and process**)** and the reaction to the students' responses. 'Modeling' includes the provision of strategies of learning or the encouragement of students to develop their own. 'Application' relates to the immediate practice of the topics taught during the lesson. 'Management of time' requires the teacher to organize his/her lesson in such a way that the students' attention spans are maximized and that they are engaged in tasks throughout the lesson. 'CLE' includes the following components: 1) teacher-student interaction, 2) student-student interaction, 3) students' treatment by the teachers, 4) competition among the students, and 5) classroom disorder. Finally, effective teachers collect information on their students' knowledge and skills in order to identify their learning needs.

The above factors cover various teaching approaches, such as constructivism and direct instruction or mastery learning. Orientation and modeling, for instance, which are the main elements of constructivism, are intended to develop students' motivation and meta-cognitive skills. In addition, via the collaboration technique, another component of the constructivism, the teacher plays an important role in making the classroom a learning environment. Furthermore, structuring and questioning are important principles in direct teaching.

The past decade, several studies have been conducted to test the validity of the dynamic model, especially in Cyprus. In the school year 2004, the findings of a longitudinal study supported the validity of the model at both the classroom level and with respect to the dimensions proposed for measuring the functioning of the separate effectiveness factors (Antoniou, 2009). Another study was a meta-analysis to estimate the effect size of the school effectiveness factors on student achievement, which provided support for the model at the school level (Antoniou, Demetriou & Kyriakides, 2006). Finally, an experimental study in which a classroom-intervention was based on this model showed an increase in teaching quality and student performance (Antoniou, ibid).

In short, EER has established that after controlling for various background characteristics, the teacher or classroom level plays a larger role than the school

level. This finding indicates the superiority of the teacher/classroom level and signifies the importance of prioritizing improvement initiatives on this plane. This section has also presented the different variables at the teacher level that affect student outcome. These variables have been found to be influenced the most by both observable and changeable teacher instructional roles. Furthermore, the discussion of EER, especially in terms of teacher effectiveness, has resulted in the presentation of a dynamic model of educational effectiveness, whose elements could be used to design teacher improvement programs. The remaining issue now is the implication of both the education standards and EER for our study, which will be explored in the following section.

2.4 Implications for the Interventions

This chapter has presented standards-based education and EER, which may serve as two relevant knowledge sources in the attainment of educational improvement. The discussion of the standards-based education described its history, its definition, and its promises. This part also discussed the criticisms and the effectiveness of the standards. Unfortunately, only a limited amount of research has been conducted to gain an understanding of the standards' effectiveness, while the existing studies have shown inconsistent, if not ineffective results. The discussion also unveiled two problems. The first one is the broadness, ambiguity, and vagueness of the standards documents. The second concerns the teachers, who have not fully benefited from the movement. Despite the fact that the role of the teachers has been widely recognized, we have concluded that this group has not received the proper support, assistance and resources.

In addition to the dynamic model, this chapter presented EER as a second knowledge-base to be used in the improvement of education. EER attempts to identify factors at different levels. In this way, it has provided a foundation which is both theory-driven and evidence-based, and which can be used to locate the aspects that should be prioritized in the educational improvement programs. The discussion in this section has shown that the teacher or classroom level plays a larger role than the higher levels, i.e. the school and regional/national levels. It is therefore important to prioritize the improvement efforts at the teacher level.

The mentioned two improvement approaches were presented as a response to educational problems, especially low student achievement, as explained in the introduction of this thesis. Standards-based education and EER share particular characteristics. For example, both are referred to in the works of Coleman, et al. (1996) and Jencks, et al. (1972), which highlights the inequity in educational opportunity and outcomes and the inadequacy of the schooling approaches to mark a starting point of influential educational movements. In view of these issues, the standards-based education has underlined the importance of setting goals or expectations in order to improve the learning opportunities and outcomes for all students, regardless of their backgrounds. Similarly, EER has defined education opportunity and high expectations as the most reliable factors in improving student performance. In addition, both the standards movement and EER, especially the dynamic model, promote the use of the constructivist approach, which prefers student engagement and cooperative student grouping to the didactic, whole-class model of instruction.

Education standards are important in the development of a shared vision on teaching and learning, which has to advocate higher expectations for all students. However, as regards the physical problem associated with the standards documents, it is essential to gain an insight into how the standards would work if they were written in more specific and concrete language. Hence, considering the findings of EER with respect to the importance of the teachers, a relevant measure would be to assist this group in improving their teaching quality, especially with respect to the factors found to be associated with student outcome. In our opinion, the combination of both approaches is likely to provide better results than the standards-based education on its own, as has been the case in Indonesia. For this reason, we developed two interventions and compared them both with a control group to test their effectiveness. The first intervention was based on the standards-based education while the second one was centered on the combination of the standards-based education and a teacher improvement program.

Prior to the interventions, two supporting documents were designed. The first one was an elaborated standards document, which was an attempt to make the existing government standards clearer, more specific, and thereby easier to measure.

The elaborated standards specifically referred to the criteria for content in relation to the performance directives in the Indonesian context. The second document included the characteristics of effective teaching, as defined by EER, especially those forming part of the classroom factors in the dynamic model (Creemers & Kyriakides, 2008).

The first intervention group was provided with the elaborated standards document and encouraged to look for effective teaching strategies themselves in order to reach the goals as defined in the standards. The second group was offered both the elaborated standards and a document containing the characteristics of effective teaching. In addition, this group also participated in a teacher improvement program, which was based on the classroom factors of the dynamic model (Creemers & Kyriakides, 2008). All eight factors of the dynamic model were introduced, while the elements of CLE and time management were combined with the aim of maximizing the students' learning opportunities.

To investigate the impact of the two interventions, the outcome variables at the student level and an intermediary variable at the teacher level were measured. Then, following EER, the background characteristics at the student, teacher, and school levels were collected to examine if they had influenced the variables, to eliminate possible bias, and to obtain more precise estimates of the intervention effects. The following figure depicts the design of the study.

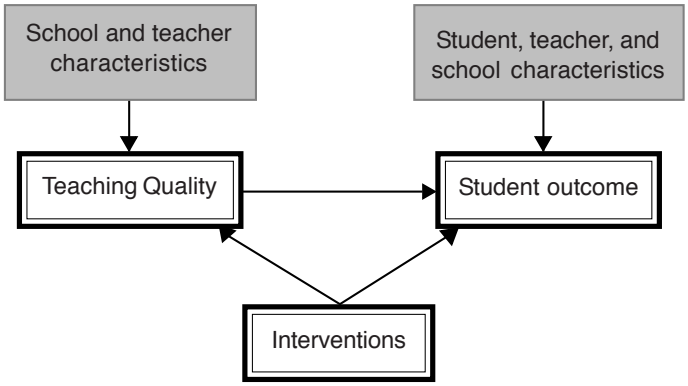


Figure 2.1 *The Design of the Study*

Based on the design and the research questions formulated, the following was hypothesized: a) Both interventions will improve the teaching quality and student outcomes, but intervention 2 will be more effective than intervention 1, and b) The effects of the interventions will be explained (mediated) by the improvement of the teaching quality. In order to test these hypotheses, one school year of experimental research was conducted, in which a longitudinal method was used, including three points of measurement. The next chapter will explain the design of the study in more detail.

CHAPTER 3

RESEARCH METHODOLOGY

3.1 Introduction

This chapter presents the research methodology used in this study. First, we will describe the research design, the research sample and the interventions. Then, the research variables and different types of instruments applied are defined in detail. The following section focuses on the research procedures; a description is given of the steps of the intervention trajectory and the data collection. Finally, the statistical analysis performed to answer the research questions will be explained.

3.2 Research Design

For this study we chose a pretest-posttest randomized experimental design, based on which the interventions were planned at predetermined times to observe their effects (Shadish, Cook, & Campbell, 2002). In addition, a longitudinal approach was integrated, according to which three measurements were set up for three groups of participants in one school year. There were two interventions: one in which a group was provided with an elaborated standards document and one in which another group was provided with a combination of the elaborated standards document and a teacher improvement program. The standards referred to the standards of content, which in the context of Indonesia was a further explanation of the standards of performance. In executing these interventions, two experimental groups and one control group were formed. This approach was used to eliminate bias and be able to estimate better cause and effect relationships (Bloom, 2006; Shadish et al. ibid; Subotnik & Walberg, 2006). The control group was offered no intervention trajectory and was asked to work solely with the government standards document. The participants voluntarily participated and were randomly assigned into the three groups as described in Table 3.1.

Table 3.1 *The Groups and the Intervention*

Group	Intervention
1. Experimental 1	The elaborated standards document; the teachers were free to develop their own strategies in implementing and achieving the standards.
2. Experimental 2	The elaborated standards document and a teacher improvement program (teacher training).
3. Control group	No intervention, teachers used the standard document available from the government.

To test the effects of the interventions, the outcome variables at the student level and an intermediary variable at the teacher level were measured. The outcome variables were cognitive (student achievement) and non-cognitive (student motivation), and the intermediary variable was teaching quality. We also collected background characteristics at the student, teacher, and school levels. This was done for three reasons: to see if these characteristics had an influence on the variables, to eliminate any possible bias, and to obtain more precise estimates of the intervention effects.

3.3 Research Sample

Although the sampling was carried out at the school level, the focus of this study was on the teacher level. As previously explained, Indonesia has a dual schooling system: general school and madrasah. General school is managed and supervised by the Ministry of National Education (MONE) and madrasah is under the authority of the Ministry of Religious Affairs (MORA). The system is divided into public and private schools. Under MONE there are more public schools while under MORA there are more private madrasahs. Both types of school are accredited every five years, with a certifications scale ranging from A (maximum) to C (minimum). The focus of this study was madrasah Tsanawiyah (MTs), a junior secondary school level (3 years after 6 years of primary schooling, age 12/13 – 14/15 years) because

in this particular type of schools improvement is urgently needed. Most madrasahs are private, small, and attended by students from low-income families. These schools generally provide education of a lower quality compared to general schools (ADB, 2006; Centre for Excellency and Quality Development Assurance [CEQDA], 2007).

Concerning the subject, we selected English because it is tested in the Indonesian national exam, and the student attainment rates on this subject have been the lowest throughout the years. Following the focus of the government in the national exam, this study has concentrated on reading comprehension. Although the other three language skills, listening, speaking, and writing are also taught, given the resources and facilities on the islands in Indonesia and their geographical barriers, it is understandable that the government has currently only focused on reading comprehension in the national exam. As regards our sample, only the second year students were involved in this study because the first year students had just started their English lessons while the third year students had to prepare for the national exam.

In terms of area, the research was limited to two neighboring provinces: DKI Jakarta and Banten. In each of these two provinces three municipalities or districts with the highest number of madrasahs were selected. Another criterion concerned school size; we selected schools with more than 100 students in total. Information on the research project and invitations to participate were sent to more than 200 schools in the selected areas.The following subsection will describe the participants and their background characteristics.

3.3.1 Participants

The plan was to organize a sample containing a minimum of 60 teachers, so that each group could consist of 20 participants. Until the end of the deadline, 57 schools (32% accredited A, 68% accredited B), including a total of 59 teachers (M = 44%, F = 56%) and 2,431 students (M = 48,5%, F = 51,5%) voluntarily participated in this study. Other schools which were invited reported different reasons not to take part in the research. Some were engaged in accreditation activities during the intervention period while others had already participated in other intervention programs. Furthermore, some teachers chose not to participate for personal

reasons. At both the teacher and the student levels, the sample population was representative in terms of gender ($X^2 = 1.01$, $df = 1$, $p = 0.31$, $X^2 = 0.02$, $df = 1$, $p = 0.88$ respectively). However, in terms of school accreditation it was not ($X^2 = 3.81$, $df = 1$, $p = .05$). In the sample there was no school accredited C, whereas in Indonesia the number of 'C-schools' is quite large.

For each municipality or district in each province the participating schools were randomly assigned into the three groups. In this way, each municipality or district had an equal chance to be in all three of them. The Serang municipality, however, formed an exception. Considering the further distance of this area from the university (UIN Jakarta), the center of our research activities, the likeliness of this group of teachers to be able to participate in the program was rather small. This is why all participating schools (N = 8) in the Serang area were positioned in the control group. As a result, the control group included more schools from Banten, whereas in the other two groups the schools from both provinces were equally divided, as shown in the following table.

Table 3.2 *The Number of Schools According to the Group and the Province*

Group	Province		Total
	DKI Jakarta	Banten	
Experimental 1	10	9	19
Experimental 2	10	10	20
Control group	6	12	18
Total	26	31	57

There were three schools (one in experimental group 2 and two in the control group) with less than 100 students. This situation could be explained by the fact that during the government's data collection these schools counted more than 100 students, whereas this number had decreased when this study was carried out. Furthermore, there was one class and one teacher in each school. The exception was two schools, in which two classes and thus two teachers in each of these schools participated. One of these schools participated in experimental group 1

and the other in experimental group 2. Because the assignment of the teachers into the groups was based on the school, the teachers from the same school were placed in the same group. In order to use all data, all four teachers were therefore included in the analysis.

As is common in a longitudinal study, the number of participants in each measurement differed, while there was a decrease in participants over time. In the first measurement the participation numbers were 1,854 students, 57 schools, and 59 teachers. By the time the second measurement took place, these numbers had decreased to 1,810 students, 50 schools and 52 classes. During this measurement, some teachers provided the wrong classes to be observed. In some of these schools, the independent observers managed to re-distribute the questionnaire and observe the right classes. As a result, the number of students did not really differ from that in the first measurement, despite the smaller number of participating schools. In the final measurement, the number of students was the least (1,691) although now there were more participating schools (54) and teachers (56) than during the second measurement. To conclude, concerning student achievement, 1,660 students (54 schools, 56 teachers) participated in both the pretest and the posttest, whereas 1,133 students (50 schools and 52 teachers) participated in all three measurements. In all these measurements, the female (both teachers and students) outnumbered the male participants.

The reasons for not participating in the second and/or the third measurement differed. Some teachers were too busy in completing their academic calendar while others were not scheduled to teach reading comprehension during the measurement period. Furthermore, a few teachers could not be contacted. They might have had second thoughts on being observed.

3.3.2 The participants' background characteristics

At the school and the teacher levels, 50 schools and 52 teachers participating in all three measurements were included in the analysis. At the student level, those who participated in both the pretest and the posttest were included in the analysis (N = 1,660). The distribution and the chi-square test served to show the differences and the similarities among the participants. Additionally, because the number of participating schools and teachers was relatively small, which was likely to render

the chi-square test insignificant, a proportion difference analysis was performed. In this case, a coefficient of .3 was used as a reference to indicate an acceptable level of difference in proportion.

At the school level, the chi-square test (of both the three groups and the paired comparison of the groups) showed no significant differences with respect to the variables. However, as Table 3.3 indicates (the bold value), the proportion difference analysis showed differences in school size, school accreditation and school score on English. Nevertheless, the majority of the schools (< 80%) were medium in size and differences in school score on English occurred only in the category of schools with a high score (8,1 – 9), which was less than 15%.

Table 3.3 *School Characteristics*

Characteristic	Distribution (%)			Difference in Distribution		
	Exp 1	Exp 2	Cont	Exp 1 and 2	Exp 1 and Cont	Exp 2 and Cont
School province						
1. DKI Jakarta	47	50	29	-.10	.10	.19
2. Banten	53	50	71	-.03	-.14	-.10
School size						
1. Small (< 100)	0	6	12	**-.33**	**-.67**	**-.33**
2. Medium (101– 500)	80	94	82	-.12	-.05	.07
3. Big (> 500)	20	0	6	**.75**	**.42**	-.25
School accreditation						
1. A	20	50	24	**-.38**	-.06	**.31**
2. B	80	50	76	.09	0	-.12
School score on the English national exam						
1. Low (5,1 – 6)	20	17	29	.00	-.18	-.18
2. Medium (6,1 – 8)	68	70	70	-.06	-.06	.00
3. High (8,1 – 9)	20	12	0	.00	**.50**	**.50**
Missing	0	5	0			

Notes. Exp 1 refers to experimental group 1, Exp 2 to experimental group 2, and Cont to the control group.

Hence, the only difference was in school accreditation, where A was considered better than B in several respects, such as school management, facilities, teacher qualification (degree), learning processes, and student performance. Experimental group 2 had an equal number of schools accredited A and B, whereas experimental group 1 and the control group included slightly more schools accredited B. With respect to the control group, the fact that eight schools in Banten had all been positioned in this group might have influenced this result.

At the teacher level, also the chi-square test did not show significant differences among the five characteristics, neither in the three groups nor in the paired comparison of the groups. The difference in proportion analysis did not indicate any difference either, except for teacher age, where experimental group 2 included a slightly larger number of younger teachers compared to the control group. Unexpectedly, this difference did not correspond with the extent of teaching experience, since neither the chi-square nor the difference in proportion analysis had displayed this difference. In brief, the participating teachers shared similar characteristics, except in terms of age.

Table 3.4 *Teacher Characteristics*

Characteristic	Distribution (%)			Difference in Proportion		
	Exp 1	Exp 2	Cont	Exp 1 and 2	Exp 1 and Cont	Exp 2 and Cont
Teacher gender						
1. Female	50	74	47	-.20	0	.20
2. Male	50	26	53	.14	-.05	-.18
Teacher degree						
1. Diploma	13	21	18	-.22	-.11	.11
2. Bachelor	81	74	77	-.03	.01	.02
3. Master	6	5	6	0	0	0
Teacher major						
1. English	75	84	82	-.1	-.05	.05
2. Non-English	19	11	18	.13	0	-.13
3. Missing/unknown	6	5	0	0	.5	.5
Teacher age						
1. <= 30 years	50	47	47	-.04	0	.04
2. 31 - 40 years	25	42	18	-.27	.07	**.33**
3. 41 - 50 years	25	11	29	.26	-.09	**-.34**
4. Missing/unknown			6			
Teaching experience						
1. <= 5 years	44	42	31	-.05	.10	.15
2. 6 - 10 years	31	32	31	-.06	0	.06
3. > 10 years	25	26	38	-.07	-.13	-.07

Notes. Exp 1 refers to experimental group 1, Exp 2 to experimental group 2, and Cont to the control group

Table 3.5 *Student Characteristics*

Characteristic	Distribution (%)			Difference in Distribution		
	Exp 1	Exp 2	Cont	Exp 1 and 2	Exp 1 and Cont	Exp 2 and Cont
Student gender						
1. Female	50	55	52	0	.07	.06
2. Male	50	45	48	.08	.11	.03
Father's education						
1. Primary	22	13	24	.22	.08	-.14
2. JSS	24	23	25	.06	.07	.01
3. SSS	33	39	26	-.02	.15	.16
4. University	6	10	4	-.17	.13	**.30**
Missing	15	16	21			
Mother's education						
1. Primary	32	19	35	.20	.07	-.12
2. JSS	23	27	19	-.03	.12	.15
3. SSS	26	28	20	.01	.15	.14
4. University	4	7	4	-.19	.08	.26
Missing	16	18	22			
Father's Job						
1.Labor and Farmer	32	26	38	.10	.03	-.07
2. Small business	40	34	38	.10	.10	.01
3. Professional	17	25	10	-.13	.18	**.31**
Missing	12	15	14			
Mother's Job						
1. Housewife	82	72	75	.09	.12	.03
2. Labor and Farmer	2	3	4	-.04	-.12	-.8
3. Small business	6	6	6	.06	.11	.05
4. Professional	5	11	4	-.27	.9	**.36**
Missing	5	9	11			

Notes. Exp 1 refers to experimental group 1, Exp 2 to experimental group 2, and Cont to the control group.

Finally, in contrast with the school and the teacher characteristics, as described in Table 3.5, the chi-square test showed significant differences among the three groups as regards all student characteristics, except for gender. The chi-square of the independent group comparisons, however, indicated that these differences occurred mostly between experimental group 2 and the control group. Similarly, as described in Table 3.5, the difference in proportion analysis generally suggested no differences, except in the paired comparison between experimental group 2 and the control group with respect to fathers graduated from university and fathers and mothers working as professionals. However, the percentages represented by these items were only small.

In sum, the above description shows that the randomization was generally successful in this study. Only small differences were observed with respect to school accreditation and teacher age, which were neither reflected by the school scores on the English national exam nor by the extent of teaching experience, respectively.

3.4 The content of the interventions

The two interventions were aimed at the improvement of student learning outcomes through enhancing the teacher-teaching quality. The literature review in the previous chapter discussed the importance of education standards in the formulation of a shared vision on teaching and learning regarding what students have to achieve. Such a vision, as indicated, was expected to improve the quality of education. However, we also learned that one of the main problems of the standards movement has been the broad and general manner in which most standards are laid down, including those of Indonesia. It was expected that specific, concrete, and measurable standards as provided in the first intervention in this study, would help especially the teachers to develop a common and concrete understanding of what to teach. In this way they could provide better learning opportunities and improve the learning outcomes of all students. However, it was also argued that learning outcomes will have a better chance of increasing when education standards and a teacher improvement program are combined. Moreover, this two-dimensional approach, specifically focused on teachers, was expected to be specifically effective considering the results of EER, as presented in chapter two. These results have shown that in any educational innovation the role of the teachers is crucial. For this

reason, the second intervention was implemented and compared to the first one. Intervention two was expected to yield better results than the first intervention and the method used by the control group.

Prior to the interventions, two supporting documents were developed: "the elaborated standards of English" and "becoming an effective teacher of English". The first document was an attempt to make the government standards more comprehensible and clear by specifically defining the reading competencies listed in the standards. The second document offered both research-based and practical strategies that teachers could use to improve their teaching quality. In addressing the reading skills defined in the elaborated standards they could use the classroom factors of the dynamic model. These two documents could be considered as "what to teach" and "how to teach" manuals, respectively, and will be further explained in sub-sections 3.4.1 and 3.4.2.

3.4.1 Intervention 1

So the first intervention group was provided with the elaborated standards document in which the reading competencies defined in the government standards were explained in more detail. The document was based on an analysis of the government standards of performance and content, a literature review on reading comprehension skills, and an analysis of the reading competencies tested in the Indonesian national exam of English at the junior secondary school level. Below a fragment of the government's performance standards document is presented. It is the standard of performance for English reading comprehension.

> **"understanding meaning** in simple written interpersonal and transactional discourses both formal and informal in the form of *recount*, *narrative*, *procedure*, *descriptive*, and *report*, in daily life context." (Pemerintah RI, 2005)

The formulation of this standard is problematic in at least two respects. Firstly, no specifications are provided for the separate grades. The reason for this may be that the competencies are only to be fully mastered when the students graduate. Nonetheless, this should be mentioned. Furthermore, although it is indicated which types of text have to be mastered, there is no indication of the number of words that students are expected to know by the time they graduate. Secondly, reading

competence is merely defined by “understanding meaning”. What this exactly entails is not explained. Standards of this kind should provide much more clarity. The terms used in the standard of content are no different, except that the grades (year) and semester (half-year) are specified. An example of the standard of content of grade VIII semester 1 is as follows: “**responding to meaning** to short and simple functional texts related to the surroundings accurately, fluently, and meaningfully”. The two prerequisites ‘understanding meaning’ (standard of performance) and ‘responding to meaning’ (standard of content) show a crucial irregularity in that the latter exceeds the first in terms of capability. It is only logical to assume that a student is capable of responding when he/she understands a meaning.

Since until now no studies have been published engaged in the clarification of these terms, the elaborated standard document was an attempt to make the existing standards more specific and concrete. This was done by adding the specific reading skills to the *understanding meaning* and *responding to meaning* items and by indicating the length of the texts for each grade. The elaborated standards document concentrated on the standards of content since they specified the grades and the semesters, which differ in terms of the length and the types of the texts provided. In order to define the reading competencies, further analysis was carried out, which included a literature review of the specific reading skills required in reading comprehension and an examination of the reading skills tested in the Indonesian national exam for junior secondary school.

Rosenshine (1980) provides observable reading comprehension skills, categorized into three general types ranging from easier (locating details) to more complex skills (inferential competencies). “Understanding meaning” refers to the easier reading skills and “responding to meaning” to the more complex competencies. The gradation of these reading skills was also reflected in the national exam: from finding factual information to predicting titles or drawing conclusions.

As an example, some specific reading skills under *understanding meaning* include recognizing words or paraphrasing meanings in the text: reading pictures, tables and numbers, as well as answering specific text-based questions. These skills are considered as basic comprehension competencies, which can be executed by reading the text. *Responding to meaning* requires specific reading skills, such

as thinking beyond the information provided in the texts, for instance extracting the main ideas from a text, making inferences, and evaluating the views presented in a text. In addition, for each reading skill assessment checklists were provided to help the teachers evaluate the students' mastery of the competencies. An example of a checklist item for recognizing words or paraphrasing meaning was the criterion of identifying the meaning of at least 80% of the words or phrases in a text.

The elaborated standards document could therefore be considered as a 'what-to teach' instruction. It was expected to provide teachers with a clearer understanding of what to plan, to teach, and to assess. It served as a reference source for the participants, who all taught the same subject in different schools, and it was meant to stimulate them in harboring the same minimum expectations for all students. Students may have different personal and academic backgrounds, which may affect the way in which they achieve the standards. However, as mandated by the standards-based education, teachers are required to help these different students achieve the same targeted goals by using the proper strategies. Together with experimental group 2, experimental group 1 attended a one-day introductory workshop to discuss the elaborated standard document in more detail.

3.4.2 Intervention 2

The second intervention contained the combination of the elaborated standards and a teacher improvement program. In this intervention an additional booklet, titled "Becoming an effective teacher of English" was used, which served as a "how to teach document". Hence, the teachers in this intervention group were not only equipped with information on "what to teach" but also with instructions about "how to teach". In addition to attending a one-day introductory workshop on the elaborated standards, this group also joined a one-day introductory workshop on effective teaching. Next, group two was provided with six half-day monthly workshops, all of which were focused on the elaborated standards and the classroom factors of the dynamic model.

The "how to teach document" discussed the characteristics of effective teaching as defined by the classroom factors in the dynamic model, linking these factors with the specific reading skills as mentioned in the elaborated standards. The document started with an example of classroom interaction between teachers and

students, followed by explanations about the meaning of these interactions. This introductory section was intended to give a simple and practical explanation of the characteristics of effective teaching on the basis of the eight classroom factors of the dynamic model.

The next section presented various teaching strategies or 'modeling' approaches. An example included in the document was semantic mapping, which has been found to be an effective strategy to build vocabularies (Grabe, 2009; Pittelman & Heimlich, 1991). This approach could be used as a model to teach students to recognize words or paraphrase meanings in a text. Furthermore, 'application', 'questioning' and 'assessment' were explained through the provision of various questions and tasks associated with the different types of reading skills as stated in the elaborated standards. The 'classroom as a learning environment' and 'time management' were combined with an emphasis on maximizing the students' learning opportunities. Finally, the document was completed with the reading instruction procedures and an example of a lesson plan. Both the elaborated standards and the effective English teaching documents are provided in the appendix of this thesis (Appendix 3.1 and 3.2).

The combination of the additional information on the reading skills instruction and the description of the classroom factors in the dynamic model was expected to improve both the quality of teaching and the students' attainment levels, as mandated by the educational standards. During the intervention period the topics covered by the elaborated standards document were divided into six half-day workshops. The first workshop was focused on orientation and structuring, aimed at helping the students become aware of the objectives of the subject and the importance of the lessons and the series of activities that had to be undertaken. Workshops two and three discussed teaching modeling and the development of application tasks and questions in relation to the reading skills *understanding meaning* and *responding to meaning* respectively. Here the teachers were strongly encouraged to introduce various strategies or modeling techniques to the students, such as semantic mapping, using a graphic organizer, and/or SQ3R (survey, question, read, recite, review). The teachers were also suggested to provide different types of feedback, especially when students were not able to answer their questions directly. With respect to reading skills *responding to meaning,* the teachers were advised to spend

more time on group work instead of whole class instruction and to vary between different text levels on the basis of student ability. It was argued that this approach would provide all students with better learning opportunities, especially because this competency of reading skills has generally been considered to be more difficult.

These three workshops were followed by lesson plan development and peer teaching. The teachers were asked to define particular learning objectives, to propose various activities as suggested by the classroom factors in the dynamic model, and to formulate a number of tasks and questions. In the last meeting, some teachers were randomly selected to demonstrate their teaching skills while others were asked to act as students or observers and to provide feedback for improvement purposes.

3.5 Research variables and instruments

As already indicated, the outcome variables in this study were student performance on English reading comprehension and student motivation to learn English, while the intermediary variable was teacher teaching quality. Student performance was measured by an English reading comprehension test while data on student motivation were collected via a questionnaire. Teaching quality was measured in two ways: through classroom observation by an independent observer using a high inference classroom observation instrument (from now on referred to as observation instrument) and via a questionnaire on students' perception of the teacher teaching quality in their classes (from now on referred to as student questionnaire). Finally, also the background characteristics of the students, the teachers and the schools were gathered through questionnaires. All instruments were translated into Bahasa Indonesia, except the English test and the classroom observation instrument.

3.5.1 Student performance on English reading comprehension

The students' performance was measured by means of a pretest and a posttest, which were developed based on the specific reading skills as described in the elaborated standards document. The reading comprehension test level A2 of the Central Institute for the Development of Tests (CITO), a leading institute on testing and assessment in the Netherlands, was used as the point of departure. It was

modified and shortened to make it match the elaborated standards and the Indonesian context.

In the two pilot studies both reliability (α) and an item difficulty analyses were performed to identify the items that should be retained in the study. The original instrument contained 46 items whereas the last pilot study resulted in 25 items with α = .78 and a difficulty level of .21 to .77. In this set, more than 50% had a difficulty level of above .60 while only two items had a difficulty level below .30. The number of 25 items was considered extensive enough for our purposes, because the students also had to complete the questionnaire on teaching quality and student motivation. These 25 items were used for both the pretest and the posttest, while two easier items (based on the pilot study) were made slightly more difficult in the posttest. Next, some examples of the items are presented.

Decide for each of the statements listed below if they are true or false, according to the text titled "Survival Course in Dutch".

The course will emphasize speaking and listening skills.

A. True
B. False

The course will be presented through video fragments.

A. True
B. False

Survival Course in Dutch

This course is intended for students having absolutely no prior ability in the Dutch language. It deals with the use of the Dutch language in everyday situations. The emphasis during the workshops will be on the development of speaking and listening skills. The matter to be studied will be presented in the form of short dialogues, which students can understand (in a general way) with the aid of texts and pictures. The duration of the course is 14 weeks: 2 hours in class each week.

Bron: www.fontys.nl

Read the following article. Do you have to book in advance if you want to check in sport equipments?

a. Yes, if we bring items which weigh more than 50kg
b. Yes, if we bring items which weigh less than 23kg
c. Yes, if we bring items which weigh between 23 – 45kg
d. None of these answers are correct

BRITISH AIRWAYS

Sporting Equipment

All customers are allowed to check in their respective free checked baggage allowance PLUS one additional item of sports equipment from the list below.

Snow and skiing equipment	**Bicycle**
Golf equipment	**Windsurfer sets**
Fishing equipment	**Hang gliders**

For supporting goods items that weigh between 23kg – 45kg (50 – 99lbs) please notify at the time of booking or at a minimum of 24 hours before departure. We will not accept oversized items exceeding 2.5m x 0.80m (98in x 39in x 31in).

Bron: *www.britishairways.com*

For the actual study, we conducted reliability, item difficulty level, and correlation analyses. It is admitted that the psychometric properties of the test in this study approached the marginal boundaries. In addition, given the outcome of the reliability analysis, only 20 items were included in the study. One of the two more difficult items in the posttest, for example, was excluded. The pretest and the posttest were significantly correlated, with a coefficient of $r = .275$, $p = .01$, whereas their reliability was modest ($\alpha = .52$) and moderate ($\alpha = .62$), respectively. In the final pilot study, however, the 20 items had a good reliability ($\alpha = .74$).

3.5.2 Student motivation to learn English

The student motivation variables included reaction or opinion when facing an exam, homework, good and bad mark, and difficulties in English. These variables were measured via an eight-item questionnaire, which was presented on a four-scale. It was developed based on the work of Hermans (1983) and a modification by Kuyper, Van der Werf, & Lubbers (2000). The reliability (α) was generally good throughout the three measurements: .67, .72 and .76, respectively. Below two examples of the items in the questionnaire are given.

1. When there is an English test, I learn ……. (than) I usually do.
   a. As hard as b. A bit harder c. Harder d. Much harder

2. When studying for English, I set …….. demands for myself.
   a. Not high b. A bit high c. High d. Very high

3.5.3 Teaching quality

As indicated earlier, we used the classroom level of the dynamic model as our framework for conceptualizing teaching quality, which was measured by using two instruments: a student questionnaire and an observation instrument. Both tools were developed based on the instruments used in the aforementioned studies in Cyprus (Creemers & Kyriakides, 2008). They were modified and reduced during the two pilots to match the Indonesian context and the specific subject of this study.

The student questionnaire

The student questionnaire included four subscales (32 items) requiring students to indicate the frequency of their activities on a five points of Likert scale ranging from *not at all* (1) to *a great deal* (5). The four subscales were 'orientation', 'instruction', 'questioning', and 'CLE', which were selected on the basis of the results of the factor and reliability (α) analyses performed in the last pilot study (α= .71 to .85). Instruction was a newly constructed subscale, which consisted of the items representing 'structuring', 'modeling', and 'application'.

However, the results of a confirmatory factor analysis (CFA) showed that the four subscales would not serve as a good model. On the other hand, an exploratory factor analysis (EFA) indicated that one scale in particular should be used in this

study, namely teaching quality, with a reliability rate in each measurement of above .85. Based on the latter factor analysis and the argument that one scale would provide a more comprehensive framework, it was decided that the main analysis in this study would be focusing on only one scale: teaching quality. Also the subscales were subjected to a reliability analysis, of which the results ranged between .62 and .82. The following table presents the subscales and examples of the items.

Table 3.6 *The Subscales and Item Examples of the Student Questionnaire*

Subscales	Example Items	N of item	α	
			M2	M3
Orientation	I understand the reason for or the importance of studying the day lesson.	4	.62	.80
Instruction	Our teacher presented tips or strategies that help us doing exercises assigned by him/her.	9	.72	.82
Questioning	When a student gave a wrong answer the teacher helped her/him to understand her/his mistake and find the correct answer.	9	.71	.80
CLE	Our teacher encouraged us to ask when there is something we do not understand.	10	.63	.70
Teaching quality	All items in the subscales	32	.86	.90

Notes. M = measurement

The observation instrument

The observation instrument consisted of 52 items and was divided into two parts. Part A included 33 items and required the observer to indicate the frequency of the activities observed on a *not at all* (1) - *a great deal* (5) Likert scale. Part B contained the remaining 19 items, and dealt with the quality of the activities observed, using a 1 (minimum) to 5 (maximum) scale. Due to the small number of participating

teachers (N < 15) in the two pilot studies, no factor and reliability analysis were carried out; the pilot studies had only been meant to adjust the instrument to the Indonesian context and the specific subject focused on in this study, which was reading. All subscales as explained by the dynamic model were used in this study, except that time management was combined with the CLE. So there were seven subscales used in the study.

The observation was conducted by a team, which consisted of five observers in total, with one observer per class observation, as determined in the study design. The team had been previously trained in using the instrument through six half-day workshops. In the first two workshops, information was provided and discussed about the design of the research, the elaborated standards, the classroom factors and the dimensions of the dynamic model. In each of the remaining workshops, the trainee observers were asked to watch a 40 minute video-taped classroom teaching/ learning session and to fill in the high inference observation instrument, which was followed by a discussion. In addition, to gain more insight into a real classroom observation, they observed a real classroom situation together. At the end of the training, the inter-rater reliability was good (generalized Kappa = .72). Depending on the observers' mobility and where they lived, each one was assigned to visit around 8 – 15 schools. Each observer visited the same school three times.

In the research, both EFA and CFA were employed after which the scales were recoded into one - three (not at all – little – enough for part A, and poor - fair – good for part B), in line with the scores of most teachers. Only on a few (N = < 10) items throughout the measurements, two or three teachers scored four. Similar to that of the student questionnaire, the CFA of the observation data showed a poor model fit, which also applied to the EFA. Therefore, it was decided to base the main analysis on one scale, the more because throughout the measurement the reliability of this scale had proven to be excellent (α= > .9). Finally, the reliability of the subscales appeared to be wide-ranged (α= .49 to .93).

Table 3.7 *The Subscales and Item Examples of the Observation Instrument*

Subscales	Example Items	N of item	α		
			M1	M2	M2
Orientation	The teacher explicitly explained the aims of reading certain topics and text and practicing specific reading skills	7	.86	.93	.90
Structuring	The teacher presented the structure of the lesson (the topic, the text, and specific reading skills) to the students.	7	.83	.89	.75
Modeling	The teacher presented the concepts or strategies that the students could use to accomplish the specific reading skills taught	6	.83	.87	.75
Application	The teacher offered the students the opportunity to use the concepts, skills or strategies that they acquired throughout the lesson or during the previous lessons.	8	.76	.72	.73
Questioning	The teacher provided useful hints when a student gave a wrong answer.	10	.80	.49	.74
Assessment	The teacher posed questions to examine what the students had understood from the day's lesson.	2	.57	.60	.69
CLE	Each student was engaged in tasks/work assigned to him/her by the teacher.	12	.54	.75	.61
Teaching quality	All items in the subscales	52	.91	.94	.94

Notes. M = measurement

3.5.4 Background characteristics

The data on background characteristics were gathered at the student, teacher, and school levels. At the student level, the variables included gender and student socio-economic status (SES), the latter of which included the father's and mother's most recent education and job. At the teacher level, the variables were gender, age, academic qualification (most recent education and major) and the scope of the teaching experience. The variables at the school level contained the status of accreditation, school size, and the mean score on the English national exam, year 2010/2011.

The data on these variables were selected through questionnaires, most of which contained multiple choice items, except for the father's and mother's job. The aim of this item was to collect information on the students' socio-economic backgrounds. The background characteristics of the teachers and the schools were collected via one questionnaire. It was distributed among the teachers either at the opening workshop or during the first school visit.

3.6. Research procedure

This sub-section explains the steps of the interventions and the data collection. The research trajectory started in mid July 2010 and was completed at the end of May 2011. All data were gathered during three measurement points: in August 2010, in January 2011, and in May 2011 as described in Table 3.8 in the following page.

The table shows that the English test was distributed twice: as a pretest and as a posttest. The classroom observations took place three times, which also applied to the questionnaire on student motivation. Due to some problems, the questionnaire on the students' perception of the teaching quality could only be administered two times. The school visits took place within the schedule of the English lessons. A visit normally took two or three hours (3 x 40 minutes).

Table 3.8 *Representation of the Intervention Steps and the Data Collection* (modified from Smith & Glass, 1987)

Introductory Workshops (July 2010)	Measurement 1 (August 2010)	Training /Monthly Workshops (1-6, September 2010 - April 2011) X1 vs X2 vs X0	Measurement 3 (May 2011)
		Measurement 2 (January 2011)	
X1 - X2	O1 - Mot1 - T1	O2 - Mot2 - Q2	O3 - Mot3 - Q3 - T2

X1	The first intervention: the use of the elaborated standards
X2	The second intervention: the use of the elaborated standards and participation in a teacher improvement program
X0	The control group: no intervention
O1, 2, 3	Observation of the teachers' teaching quality by an independent observer during three measurements
Mot1, 2, 3	Student questionnaire measuring student motivation during three measurements
T1, 2	English pretest and posttest
Q2, 3	Student questionnaire measuring student perception of the teachers' teaching quality during measurement 2 and 3

3.6.1 The introductory workshops

The introductory workshops, a one day session focused on the elaborated standards for both experimental groups and another day during which the effective teaching program was introduced to experimental group 2, were conducted on July 16 and 17, 2010, respectively. There were important reasons for conducting these workshops before the first data collection. Firstly, the schools had just started, so

the participants' teaching and administrative work loads were not yet so high that they could not attend the workshops. Secondly, there was a two-week holiday at the end of August until the first week of September in connection with the fasting month and the celebration of Idul Fitri 2010[1]. So the only possibility to plan the workshops after the first data collection would have been somewhat later in September, when the teaching loads had already increased, especially as it was common for teachers in the private schools to work at several locations.

Because of these reasons and the knowledge that teachers would need a considerable amount of time to change their ways of teaching, it was considered only logical to assume that they would not have already adjusted their methods immediately after the workshops when the first data were gathered. Moreover, not all participants came to the workshops. During the introductory workshops, also the teacher questionnaire for gathering teacher and school characteristics was distributed.

Concerning the number of participants, 33 of the 41 invited teachers confirmed their attendance to the first workshop, which was intended for both experimental groups. Others could not participate because of other meetings at their schools which they had not been able to cancel. However, only 26 (10 from experimental group 1 and 16 from experimental group 2) actually came to the workshop. A number of teachers indicated they were sick that day while others decided not to come due to the bad weather and the fact that they lived relatively far from the university, where the workshop took place. The second workshop was attended by all 16 teachers from the experimental group 2, who had also participated the first day.

Most of these teachers confirmed that they taught grade 8, which was the target group in the study. In addition, they also taught in other grades. However, although having been sent to the workshop by their principals, one teacher from experimental group 1 and two teachers from experimental group 2 had no fixed assignments as regards the grade they were required to teach. It appeared later that they were not assigned to teach grade 8. We then invited all teachers of experimental group 1 who had been absent in the introductory workshop, including the newly assigned

[1] This holiday was related to a Muslim festivity. Although Indonesia is not Islamic, it is one of the most Muslim-populated countries in the world.

teacher, to spare some time after the first measurement to discuss the content of the elaborated standards. This activity took place straight after the measurement. And on September 25, 2010, when the first half-day monthly meeting was held, we also invited the teachers of experimental group 2, including the two newly assigned ones, to come in three hours earlier to discuss the materials presented in the introductory workshop.

3.6.2 Measurement one

Measurement one was conducted on August 2010, when four activities were carried out: classroom observation, distributing the questionnaire on student motivation, administering the pretest, and allocating the teacher questionnaire to gather data on the teacher and school characteristics. During the first hour of the lesson, for about 30 – 40 minutes, the classroom observation took place. The next step was the distribution of the student motivation questionnaire, which took around 5 minutes. Via this questionnaire also student background characteristics were collected. The final step was administering the pretest, which took around 45 minutes. The teacher questionnaire was distributed when the students did the test. In intervention 2 (based on the combined approach), the observer provided the teachers with feedback after the lesson.

3.6.3 Training / monthly workshops for experimental group two

The monthly workshops consisted of a combination of brainstorming, lecture, discussion and group work. As regards this last component, the participants presented their results. A senior lecturer of the English Department of the University of UIN Jakarta assisted in facilitating these meetings. During the intervention period there were six half-day workshops, held from September 2010 – April 2011. During the meetings, different types of texts and publications from various sources, including textbooks, were studied in groups. In addition, after each workshop the participants received feedback.

With respect to the attendance rates, not all 21 teachers of experimental group 2 visited the workshops, and those who did, did not join all six meetings. At the first two workshops, 16 and 15 teachers, respectively, were present while the remaining four workshops were attended by 10 teachers on average. Nevertheless, there were

no teachers who stopped their participation or changed grade. Most of them attended four workshops in total while four teachers even visited all six of them. The reasons of those who could not participate varied, for example being ill, having to attend family affairs or being engaged in school-related activities.

3.6.4 Measurement two

Measurement two took place in the middle of the training period (January 2011). In addition to the classroom observation and the distribution of the questionnaire on student motivation, also the questionnaire on the students' perception of the teaching quality was handed out. Similar to measurement one, the classroom observation took place during the first hour of the lesson. After that the student questionnaires on motivation and teaching quality were distributed, the latter of which took around 15 – 20 minutes. Per student the two questionnaires on student motivation and background characteristics were put together so we could locate the data provided by each of them in each measurement.

3.6.5 Measurement three

The last measurement took place in May 2011, after the training session. After the classroom observation, the questionnaires on student perception and teaching quality were handed out. The final activity was administering the posttest, which took place eight months after the pretest (August 2010).

3.7. Data analysis

For assessing the data two approaches were used: descriptive and multilevel analysis. First, all data were subjected to a reliability (á) analysis, after which those supplied by the independent observers and those on the students' perception of the teaching quality were measured by performing a factor analysis. Next, the background characteristics of the participants (schools, teachers, and students) were analyzed to check whether the random assignment had been successful.

The further analysis focused on the descriptive statistics of the intermediary and the outcome variables. Concerning the intermediary variable, both teaching quality and its subscales were analyzed. This was done to make the changes throughout the measurements visible. While the main analysis particularly

concentrated on the effects of the interventions, the descriptive analysis was concerned with the differences among the three groups as regards the intermediary and the outcome variables. Here descriptive statistics were performed as well as analyses of significance and of effect size (Cohen's *d*). In order to gain an insight into the underlying factors that had contributed to the differences among the groups, the relationship (*r*) between the different characteristics and the intermediary and the outcome variables was examined. Here the coefficient of $d = .2$ was considered as small, $d = .5$ as medium, and $d = .8$ as large, and that of $r = .10$ as small, $r = .30$ as medium, and $r = .50$ as large (Cohen, 1992). In this study, $r = .20$ was considered as the minimum value of a correlation. To obtain better estimates of the intervention effects, the results of these analyses were used to decide which background characteristics should be included as explanatory variables in the multilevel analysis.

The multilevel analysis to test the effects of the interventions (Goldstein, 2003; Snijders & Bosker, 1999) was conducted in MLwiN (Rasbash, Charlton, Browne, Healy, & Cameton, 2005). The multilevel approach was used because of the hierarchical structure of the data. In the analysis, dummy variables were set up for both interventions by positioning the other groups as the control group. Following the main research question, the first analysis examined the effects of the two interventions on teaching quality. The next multilevel analysis tested the effects of the interventions on student outcome: the students' performance in English reading comprehension and their motivation to learn English. In analyzing student achievement, both covariance and learning gain analyses were employed. Finally, the last analysis investigated whether the improvement of teaching quality explained the student outcomes. In each analysis, several models were built in addition to the empty model. Chapter 5 presents the details of the multilevel analysis.

For the analyses the following significance levels were accepted: $p < .01$ for the data at the student level and $p < .05$ for the data at the teacher or school level because of the small number of participants. In general, the tests were two-tailed, except when examining the differences among the three groups after the first measurement as regards the intermediary and the outcome variables in the descriptive part. It was expected that differences in teaching quality, student achievement, and student motivation would only be found in the second and the

third measurement, which is why here one tail was considered appropriate. A one-tailed test was also used in estimating the effects of the interventions on the intermediary and the outcome variables and in measuring the relationship between the improvement in teaching quality and the improvement in student outcome. For the hypothesis as regards the effects of the interventions on the relationship between the improvement in teaching quality and the enhancement in student outcome, we also performed a one-tailed test. In the multilevel analysis, the *t*-ratio coefficient was used to test the significance of the fixed effects of the models, with a *t*-value which had to be higher than 1.96 for $p < .05$ (Snijders & Bosker, 1999).

CHAPTER 4

DIFFERENCES AMONG THE THREE GROUPS WITH RESPECT TO THE INTERMEDIARY AND THE OUTCOME VARIABLES; Descriptive Findings

4.1 Introduction

The main objective of this chapter is to present the descriptive findings of the study, which concern the differences among the three groups with respect to the intermediary and the outcome variables. The chapter will describe the groups' mean scores, standard deviations and effect sizes (*d*) for teaching quality, student achievement, and student motivation. In addition, we report the results of the variance analysis (ANOVA) and the *t*-test, which examined the differences in and among the three groups as regards the above three variables.

Also the relationship (*r*) between the background characteristics at the student, teacher, and school levels and the three variables will be explained. It was considered important to understand which characteristics played a role in these differences and had therefore to be retained in the multilevel analysis to provide better estimates of the intervention effects. The intervention effects on these variables were tested in the multilevel analysis, which are presented in the next chapter.

First, the differences in teaching quality, student achievement, and student motivation are described, followed by the relationships between the background characteristics and teaching quality and student outcome. After that, we present the descriptive findings and list the background characteristics which had to be retained in the multilevel analysis.

4.2 Differences among the Groups with respect to Teaching Quality and Student Outcome

This section maps out the differences among the three groups with respect to teaching quality and student outcome. As regards teaching quality, both the observer and the student data will be reported. As previously mentioned, the observers conducted three measurements, while the students rated their teachers' teaching quality only two times, namely during measurement two and three. Furthermore, the subscales of teaching quality will be listed; seven for the observer data and four for the student data. Finally, the last two sections will deal with the differences in student achievement and student motivation.

4.2.1 Teaching Quality

This subsection provides a summary of the main characteristics of the data, where teaching quality was the sum of all items divided by the total number of the items. There were no missing values in the case of the observer data, while also almost all students completed all items. As Table 4.1 shows, the mean scores and the standard deviation of teaching quality in each group in each measurement are described and calculated by the effect size analysis (Cohen's *d*), and their significance is reported from measurement to measurement. Furthermore, graphs have been added of the mean scores of teaching quality of the three groups in all three measurements to provide a better understanding of how each group developed throughout these sessions.

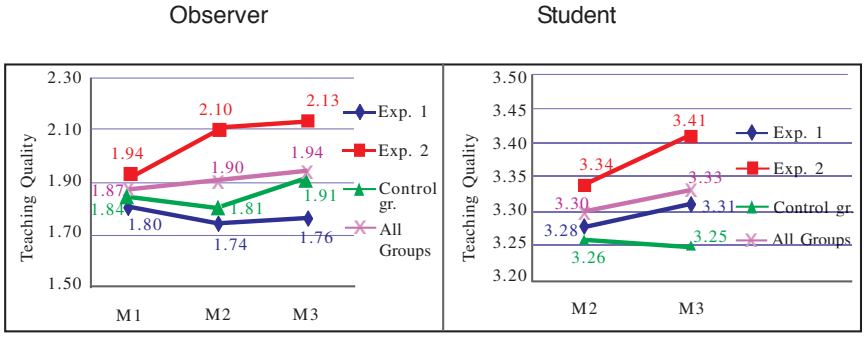
First of all, the scales of the observation instrument were reduced from one to five to one to three because of the small number of teachers who scored four and five. Based on the original scale of the observation instrument, as reported in the above table, the teachers were rated much higher by the students than by the independent observers. This result can be explained by the Indonesian culture, where the teaching profession is commonly regarded as a high and respecful occupation. Teachers therefore enjoy the respect of society in general and of students in particular (Maulana, Opdenakker, Den Brok, & Bosker, 2010). However, their patterns were similar, which will be explored later in this section.

Table 4.1 *The Mean Score, the Standard Deviation, and the Effect Size of Teaching Quality Based on the Observer and Student Data collected during Three Measurements*

Group	M 1		M 2		M 3		Effect size (*d*)		
	Mean	*SD*	Mean	*SD*	Mean	*SD*	M1 to M2	M2 to M3	M1 to M3
Observer data									
Exp. 1	1.80	.26	1.74	.36	1.76	.22	-.19	.07	-.17
Exp. 2	1.94	.37	2.10	.34	2.13	.35	.44*	.09	.53**
Cont.	1.84	.34	1.82	.30	1.91	.23	-.06	.33	.24
Total	1.87	.31	1.90	.37	1.94	.31	.09	.12	.22
Anova	*F*(2, 98) = .98		*F*(2, .611)= 5.34**		*F*(2, .84)= 3.74**				
Bonferroni test (mean difference (*MD*) / SD)									
Exp 1 & Exp 2	-.14(.11)		-.35(.11)*		-.37(.09)*				
Exp 1 & Cont.	-.04(.11)		-.08(12)		-.15(.10)				
Exp 2 & Cont.	.10(.10)		.27(11)*		-.22(.09)*				
Student Data									
Exp. 1			3.28	.48	3.31	.55		.06	
Exp. 2			3.34	.48	3.41	.56		.13**	
Cont.			3.26	.45	3.26	.56		-.02	
Total			3.30	.47	3.33	.56		.06	
Anova			*F*(2, .84) = 3.77		*F*(2, 2.80) = 8.92**				
Bonferroni test (mean difference (*MD*) /SD)									
Exp 1 & Exp 2			-.06(.03)		-.10(.04)**				
Exp 1 & Cont.			-.02(.04)		.06(.04)				
Exp 2 & Cont.			.09(.04)		.16(.04)**				

Notes. M = measurement, Exp 1 = experimental 1, Exp 2 = experimental 2, Cont = control group, * $p < .05$ (1 tailed), ** $p < .01$ (1 tailed)

At this point, it is important to note that for measurement one, the significance test of the observer data showed no differences in teaching quality, neither among the three groups nor among comparisons (Bonferroni test) of the groups. As expected, significant differences occurred in measurement two and three. As regards these two measurements, the Bonferroni test marked significant differences in the comparisons between experimental group 2 and experimental group 1, which also applied to experimental group 2 and the control group. Rather different, students noticed significant differences in the three groups only in measurement three. During this measurement, it appeared that, similar to the observers, the students also recognized significant differences between experimental group 2 and experimental group 1 as well as between experimental group 2 and the control group. Table 4.1 Indicates the sizes of these differences.



Notes. M = measurement, Exp 1 = experimental 1, Exp 2 = experimental 2, Control gr. = control group

Figure 4.1 *The Pattern of Teaching Quality Based on Observer and Student Data (mean score)*

Referring back to the patterns of judgment, the above figure clearly shows that the general trend of teaching quality from the points of view of both the observers and the students indicated similar trajectories: an overall positive direction of improvement. In addition, Figure 4.1 clearly depicts that the observers recognized a positive significant improvement of experimental group 2, although it has to be added that the improvement from measurement two to three was not significant. On the other hand, although not significant, the other two groups indicated a decrease in teaching quality from measurement one to two, whereas in measurement three the improvement curve rose again. The figure also shows that the students in experimental group 2 perceived a larger improvement than the students in the control group and experimental group 1, depicted by a steeper line. Furthermore, in both measurement two and three, the judgments of the students and the independent observers were significantly correlated, with coefficients of $r = .44$, $p = .00$, and $r = .35$, $p = .02$, respectively.

Related to the above findings, the effect sizes in the right side of Table 4.1clearly shows that, according to the observers, only experimental group 2 improved substantially ($d = .44$) between measurement one and two, which improvement was significant. However, the improvement of this group from measurement two to three was only very small ($d = .09$) and not significant. Similarly, the students in experimental group 2 also recognized a small but significant improvement between measurement two and three ($d = .13$). The results of the significance test are provided in the appendix (see Appendix 4.1).

4.2.2 The Subscales of Teaching Quality

As already mentioned, the observer data contained seven subscales, based on those of the dynamic model, except that 'time management' was combined with 'the creation of the classroom as a learning environment' (CLE). These subscales were 1) orientation, 2) structuring, 3) modeling, 4) application, 5) questioning, 6) assessment, and 7) CLE. The subscales of the student data were based on the factor analysis calculation in the pilot study, which resulted in only four subscales: 1) orientation, 2) instruction, 3) questioning, and 4) CLE. Instruction was a new scale, containing the items structuring, modeling, and application.

Table 4.2 *The Mean Scores of the Subscales of Teaching Quality Based on the Observer Data*

Group	Orienta-tion	Structu-Ring	Modeling	Applica-tion	Questio-ning	Assess-ment	CLE
Measurement one							
Experimental 1	1.52	1.79	1.48	1.62	2.13	2.00	2.08
Experimental 2	1.86	1.92	1.57	1.60	2.29	2.21	2.28
Control group	1.62	1.69	1.59	1.47	2.27	2.08	2.21
Total	1.68	1.80	1.55	1.56	2.23	2.10	2.19
Measurement two							
Experimental 1	1.73	1.70	1.47	1.41	1.66	2.00	2.26
Experimental 2	2.30	2.21	1.80	1.72	1.87	2.55	2.51
Control group	1.65	1.64	1.54	1.57	1.78	2.05	2.39
Total	1.91	1.87	1.61	1.57	1.82	2.22	2.40
Measurement three							
Experimental 1	1.71	1.78	1.52	1.45	1.68	2.03	1.97
Experimental 2	2.30	2.12	1.87	1.83	1.98	2.50	2.26
Control group	1.74	1.83	1.71	1.66	1.97	2.17	2.33
Total	1.93	1.92	1.70	1.65	1.82	2.25	2.19

The above table tells us several points. Of the seven subscales, the observers scored higher on the three subscales questioning, assessment, and CLE, which means that they were more positive on these. This finding suggests that in general the teachers started to use these factors in their teaching: questioning to organize their classroom activities, assessment to check the students' level of understanding, and CLE to adjust the classroom to facilitate learning. Furthermore, as Table 4.3 indicates, although the student data show a smaller degree of variation in their judgments as regards the different subscales (mean score = 3.21 – 3.48), the data do reveal that they were also positive about 'questioning' than other subscales, as this is one of the two subscales rated higher by the students. The same applies to 'orientation'.

Table 4.3 *The Scales of Teaching Quality Based on the Student Data*

Group	Orientation		Instruction		Questioning		CLE	
	M2	M3	M2	M3	M2	M3	M2	M3
Exp. 1	3.45	3.58	3.21	3.32	3.39	3.36	3.22	3.30
Exp. 2	3.53	3.62	3.24	3.37	3.46	3.43	3.30	3.41
Cont.	3.46	3.44	3.16	3.31	3.39	3.24	3.16	3.20
Total	3.48	3.56	3.21	3.33	3.42	3.36	3.23	3.32

Notes. Exp. 1 = experimental 1, Exp. 2 = experimental 2, Cont. = control group, M = measurement

On the other hand, the observers were negative in their judgments with respect to the subscales orientation, structuring, modeling, and application, which means that as regards orientation they disagreed with the students. Nevertheless, both groups agreed upon all the items of the subscales structuring, modeling, and application. The students also rated 'instruction' (containing the items structuring, modeling, and application) lower than the other ones. These findings suggest that structuring, modeling, and application were no easy activities for the teachers. Furthermore, the observers and the students also disagreed upon orientation and CLE. Whereas the students were positive about orientation and negative about CLE, the observers were negative about orientation and positive about CLE. Thus, there was agreement on all subscales except for orientation and CLE.

Table 4.4 *The Effect Size of the Subscales of Teaching Quality Based on the Observer Data*

Sub-scale	Effect size (*d*)								
	Experimental 1			Experimental 2			Control group		
	M1 to M2	M2 to M3	M1 to M3	M1 to M2	M2 to M3	M1 to M3	M1 to M2	M2 to M3	M1 to M3
Orientation	**.43**	-.04	**.42**	**.77**	0	**.77**	.08	.27	**.38**
Structuring	-.18	.19	-.02	**.55**	-.17	**.39**	-.11	**.50**	**.43**
Modeling	-.03	.23	.18	**.52**	.20	**.73**	-.13	**.51**	**.35**
Application	**-.52**	.15	**-.41**	.28	.27	**.52**	.26	**.32**	**.57**
Questioning	**-1.66**	.08	**-1.57**	**-1.43**	**.36**	**-.85**	**-1.30**	-.04	**-1.28**
Assessment	0	.05	.05	**.68**	-.09	**.61**	**-.37**	.23	.23
CLE	**.66**	-.25	**.58**	**1.06**	0	**1.06**	**.96**	.23	**1.33**

Notes. M = measurement

Table 4.5 *The Effect Size of the Subscales of Teaching Quality Based on the Student Data*

Group	Effect size (*d*), measurement 2 to 3			
	Orientation	Instruction	Questioning	CLE
Experimental 1	.20	.17	.05	.14
Experimental 2	.13	.20	.05	.20
Control group	-.03	.21	-.23	.10

As regards the groups' subscale curves, the above two tables representing the effect sizes suggest that they varied among the three groups, among the subscales, and among the measurement points. As Table 4.4 shows, with respect to orientation and CLE the observers considered the three groups to be achieving small to substantial improvements from measurement one to measurement three. The students in the experimental groups also recognized the improvement as regards

these two subscales but the students in the control group did not agree upon the improvement of their teachers in terms of orientation (d = -.03).

Another interesting result is the decreasing line for questioning in all three groups as indicated by the observer data, especially from measurement one to measurement two, which was supported by the student data. In the control group the students saw a decrease in performance as regards questioning (d = -.23). In both experimental group 1 (d = .05) and experimental group 2 (d = .05), however, the conditions in this respect were perceived as rather stable. It is nevertheless important to emphasize that in experimental group 2 the decreasing line for questioning was the least steep. Furthermore, the observers concluded that experimental group 2 managed to improve on all subscales, except for questioning, whereas the other groups showed declining rates for three subscales.

4.2.3 Student Achievement

The analysis of student achievement included 1,600 students, who participated in both the pretest and the posttest. The score was the sum of the correct answers. As Table 4.6 indicates, the Anova test showed significant pretest differences among the three groups, suggesting differences in student achievement from the beginning of the interventions. In the pre-test both experimental groups started off with significantly higher results than the control group, whereas no difference was found in the comparison between experimental group 1 and experimental group 2. Considering the students' SES, as explained in the previous chapter, this finding might have been influenced by a small percentage of parents in both experimental groups, who were higher educated and had better jobs.

Table 4. 6 *Student Achievement in the Pretest and the Posttest*

Group	Pretest		Postest		Effect Size (*d*)	*t*-test
	Mean	*SD*	Mean	*SD*		Pre – Post
Experimental 1	9.77	2.88	9.55	2.99	- .07	*t* (624) = 1.48
Experimental 2	9.92	3.14	10.65	3.47	.21	*t* (553) = -4.44**
Control group	8.78	2.98	8.13	3.34	-.20	*t* (480) = 3.65**
Total	9.54	3.04	9.50	3.40	-.01	*t* (1659) = .32
Anova	*F*(2, 194.82)=21.60**		*F*(2, 817.50)=72.02**			
Bonferoni test (mean difference /SD)						
Exp 1 and Exp 2	-.15(.17)		-1.10(.19)**			
Exp 1 and Cont.	.99(.18)**		1.42(.20)**			
Exp 2 and Cont.	1.13(.19)**		2.52(.20)**			

Notes. Exp. 1 = experimental 1, Exp. 2 = experimental 2, Cont = control group, **= $p < .01$ (two-tailed)

Furthermore, also as regards the posttest the Anova test showed significant differences among the three groups. And whereas according to the Bonferroni tests no difference was found in the comparison between experimental group 1 and experimental group 2 in the pretest, significant differences were observed in all comparisons in the posttest, which indicated that the difference had become larger. At the end of the intervention, experimental group 1 maintained its higher achievement rate compared to the control group, while experimental group 2 scored the highest, compared to both the control group and experimental group 1.

Unexpectedly, the total mean score of the three groups remained stable at the end of the year, as shown by the paired *t*-test in Table 4.6, which suggests that there was no improvement. Similarly, the effect sizes of the three groups indicated neither an improvement nor a decrease. However, considering both the significance test and the effect size, experimental group 2 had made a significant improvement. On the other hand, experimental group 1 showed a decrease, although not significant. Also the control group marked a significant decrease, as shown by both the effect size and the significance test.

4.2.4 Student Motivation

Student motivation was analyzed based on the number of students participating in measurement one, two, and three (N = 1333). Similar to teaching quality, student motivation was calculated as the sum of all items divided by the total number of items. Almost all students completed all items. Table 4.7 suggests that at the beginning of the intervention, the students' motivation significantly differed per group. The Bonferoni test revealed these differences in the comparison between experimental group 1 and the control group, and between experimental group 2 and the control group. No significant difference was found in the comparison between experimental group 1 and 2.

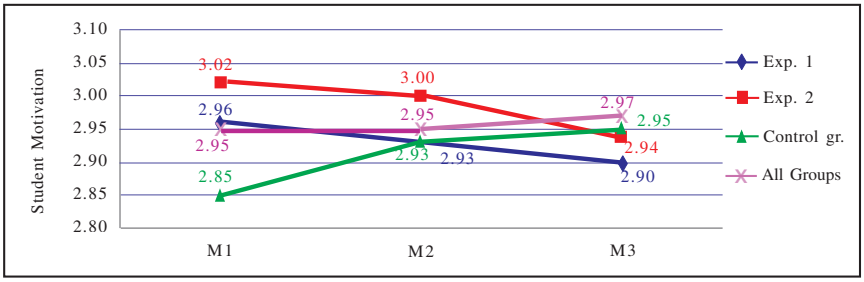
Table 4.7 *Student Motivation in the Three Measurements*

Group	M 1		M 2		M 3		Effect size		
	Mean	SD	Mean	SD	Mean	SD	M1 to M2	M2 to M3	M1 to M3
Exp. 1	2.96	0.44	2.93	0.43	2.90	0.46	-.07	-.07	-.13
Exp. 2	3.02	0.44	3.00	0.48	2.94	0.43	-.04	-.13	-.18**
Control	2.85	0.44	2.93	0.43	2.95	0.47	.18**	.04	.22**
Total	2.95	0.45	2.95	0.45	2.93	0.45	0	-.04	-.04
Anova	*F* (2, 2.63) = 13. 57**		*F* (2, .54)= 2.63		*F* (2, .24) = 1.20				
Bonferoni test (mean difference /SD)									
Exp. 1 & Exp. 2	-.06(.03)		-.06(.03)		-.03(.03)				
Exp. 1 & Cont.	.11(.03)**		.01(.03)		-.05(.03)				
Exp 2 & Cont.	.17(.03)**		.07(.03)		-.01(.03)				

Notes. M = measurement, Exp. 1 = experimental 1, Exp.2 = experimental 2, Cont. = control group, ** $p < .01$

However, the Anova test yielded insignificant results in measurement two and three. None of the group comparisons resulted from the Bonferoni tests in these measurements suggested significant differences, which suggests that the differences in student motivation gradually disappeared throughout the school year. In line with these findings, it is not very surprising that the general trend of motivation

among the three groups moved into a negative direction, although not significant, as shown in figure 2. Both experimental groups show a continuous decrease in student motivation from measurement to measurement, where the decrease shown by experimental group 2 was significant from measurement one to measurement three (see Appendix 4.2). Surprisingly, the control group managed to develop a linear pattern of improvement, although from measurement two to measurement three the increase was not significant. In sum, from measurement one to measurement three, the control group achieved an improvement of $d = .22$, whereas the rates of experimental group 1 and 2 declined to $d = -.13$ and $d = -.90$, respectively.



Notes. M = measurement, Exp 1 = experimental one, Exp 2 = experimental two

Figure 4.2 *The Pattern of Student Motivation in Three Measurements*

However, considering the variation in student motivation and student achievement, a relevant question was whether the two were related. The correlation analysis showed that although to a very small extent, motivation was positively correlated with student outcome in both the pretest and the posttest ($r = .11$, $p = .00$ and $r = .12$, $p = .00$, respectively).

In short, as regards both teaching quality and student achievement the descriptive analysis performed at the end of the intervention identified significant differences between experimental group 2 and experimental group 1 as well as between experimental group 2 and the control group. In other words, experimental group 2 demonstrated a higher level of teaching quality and student achievement than the other two groups at the end of the school year. However, in all three groups

the students' motivation had not changed by the end of the intervention. This result might have been influenced by a significant decrease in motivation in the experimental group from measurement one to measurement three. Whether the results regarding teaching quality and student outcome as presented by the multilevel analysis correspond with those of the descriptive analysis will be investigated in the next chapter. First, in the next section we will present the relationship between the different characteristics and these two items. This analysis was important to decide which characteristics were to be retained as covariates in the multilevel analysis to obtain better estimates of the intervention effects.

4.3 The Relationships between the Background Characteristics and Teaching Quality and Student Outcomes

This section deals with the question to what extent the characteristics of the participants were related to the intermediary variable and the outcome variables. In answering this question we present the Pearson correlation coefficients (*r*) between the school and the teacher characteristics and teaching quality and between the school, the teacher, and the student characteristics and student achievement and student motivation. In addition, significance tests (*t*-test and Anova) were employed to examine whether the individual characteristics made a difference (see the appendix). The results of the correlation were calculated based on the mean score of each variable throughout the measurement trajectory so that all measurement points were accounted for.

4.3.1 The Relationships between School and Teacher Characteristics and Teaching Quality

The data were analyzed based on the observer data, since this information covered all three measurement points. Fifty two teachers were assessed by independent observers. Table 4.8 clearly shows that the correlation coefficients between both the school and the teacher characteristics and teaching quality were all small. In addition, they were all not significant. The results of the significance tests showed a similar pattern; also they did not indicate any difference in teaching quality as a result of school and/or teacher characteristics (see Appendix 4.3).

Table 4.8 *The Correlations between School and Teacher Characteristics and Teaching Quality*

School Characteristic	*r*	Teacher Characteristics	*r*
Province	.10	Gender	-.19
School size	-.15	Degree	.08
School Accreditation	-.10	Major	-.02
School national exam	.12	Age	.-.11
		The degree of teaching experience	-.05

At the school level, school size had the highest (negative) correlation with teaching quality ($r = -.15$), which suggests that the larger the size of the school, the lesser the quality of the teachers. At the teacher level, gender showed the highest correlation ($r = -.19$), which means that as regards teaching quality males seemed to be lesser teachers than females. As mentioned in chapter 3, however, there was a small difference in the proportion of teachers' age. Furthermore, the significance analysis did not identify any differences in teaching quality, neither due to age nor in terms of other characteristics at the school and/or the teacher levels. Considering these results, therefore, neither school nor teacher characteristics had an influence on teaching quality.

4.3.2 The Relationship between School, Teacher, and Student Characteristics and Student Outcome

As already indicated, the analysis was based on the number of students who had participated in both the pretest and the posttest. Table 4.9 in the following page describes the correlations between the characteristics at the school, teacher, and student levels and the student outcomes. With respect to student achievement, all characteristics on the school and the student levels showed significant correlations, except the mother's job at the student level, which might have been caused by the small sample size. Similarly, the significance tests (see Appendix 4.4 for the results) indicated that all school and student characteristics significantly influenced student achievement. At the teacher level, three characteristics were found to be significantly related to student achievement: teacher degree, teacher major, and years of teaching

experience. In addition, the significance analysis revealed a significant relationship between student achievement and teacher age.

It should be noted, however, that these correlations were small. At the school level, only 'province' and 'score on the English national exam' reached an acceptable level of correlation (r = -.20). These findings suggest that the student achievement levels of the schools in Jakarta and those with a high score on the English national exam were higher. At the teacher level, the correlations of all characteristics were below the acceptable rate, although almost all characteristics made significant differences. At the student level, both the father's education (r = .22) and the father's job (r = .20) showed sufficient correlations with student achievement, implying that the higher the education and the better the job of the father, the higher the level of student achievement.

Table 4.9 *The Correlations between School, Teacher, and Student Characteristics and Student Achievement and Student Motivation*

Characteristics	Achievement	Motivation
School characteristics		
Province	-.20**	-.06
School size	.13**	.03
School accreditation	-.17**	-.01
School national exam	.20**	.13**
Teacher characteristics		
Teacher gender	-.04	-.00
Teacher degree	.10**	.13**
Teacher major	-.14**	.05
Teacher age	.04	.08**
The degree of teaching experience	.12**	.08**
Student characteristics		
Student gender	.09**	.10**
Father education	.22**	.10**
Mother education	.19**	.06
Father job	.20**	.03
Mother job	.06	-.02

** p <. 01 (two tailed)

Concerning student motivation, the only school characteristic with a significant correlation was 'school score on the national exam'. Similarly, the significance test showed that only the national school exam significantly influenced student motivation. At the teacher level, although small, 'teacher degree', 'teacher age', and 'teacher teaching experience' were significantly related to student motivation. In line with these findings, the significance test also indicated significant differences resulting from these three characteristics. At the student level, only gender and father's education showed significant correlations and made significant differences. However, it should be noted that the correlation coefficients at the school, the teacher, and the student levels were all below the acceptability threshold.

4.4 Summary and Preliminary Conclusions about the Descriptive Findings

This chapter has presented the descriptive findings of the study. The differences of the three groups in teaching quality, student achievement, and student motivation were investigated by means of a significance analysis and effect size calculations. Furthermore, we examined the background characteristics which might have contributed to the differences in the above three variables through a Pearson (*r*) correlation analysis performed at the school, teacher, and student levels. The findings are summarized in the following sub-sections.

4.4.1 Differences among the three groups

The results yielded by both analyses (that of difference and that of effect size) confirm the impact of the second intervention on teaching quality: only a significant improvement over time was detected in experimental group 2. However, this improvement only occurred between measurement one and two, after which no further improvement was noticed. This effect was recognized by both the observers and the students; in both measurements two and three, their judgments moderately correlated. Given this agreement, only one data source was used in examining the effects of the interventions on teaching quality in the multilevel analysis. For this purpose we preferred the observers' data for several reasons. Firstly, the observers were trained in conducting assessments, and hence they had more knowledge and professional background in judging teaching quality. Secondly, they observed the teachers three times, whereas the students did only twice. Therefore, the observer

data provided a more comprehensive picture of the teachers' teaching quality. Thirdly, the observers were considered to be a more independent party than the students because the latter knew the teachers personally and had dealings with them on a day-to-day basis. The observers, however, did not know the teachers. They only came to visit the schools during the intervention, which made them much more objective than the students.

We also found that the second intervention had an effect on student achievement. The three groups together showed significant pretest/postest differences, but with higher scores for the experimental groups. However, whereas no pretest differences were established between experimental group 1 and experimental group 2, we did find a posttest difference between them, namely a higher score for the second intervention group. In addition, the results of the paired *t*-test (calculation of the pattern of development between the pre- and the posttests), which was done for each group, showed that only experimental group 2 had made a significant improvement.

Student motivation, which had been there in different degrees for all three groups at the beginning of the school year, gradually disappeared throughout the interventions. Both experimental groups showed declining patterns, of which that of experimental group 2 was significant. Hence, this effect appeared to be negative. Surprisingly, a significant improvement was observed for the control group. In sum, only the second intervention seemed to have significant effects, a positive influence on teaching quality and student achievement and a negative impact on student motivation.

4.4.2 Background characteristics retained for the multilevel analysis

The next question was which characteristics might be related to the differences among the three groups and should therefore be retained in the multilevel analysis as covariates. Four aspects were considered in their selection: 1) their distribution at the school, teacher and student levels, 2) their correlations with teaching quality, student achievement, and student outcome, 3) the extent to which they made a difference, as indicated by the significance test, and 4) the extent to which the literature review supported the correlations between the characteristics and the three variables. The correlation (*r*), for that matter, had to be at least .20.

In terms of teaching quality, neither the school nor the teacher characteristics showed correlations strong enough to make any difference. This finding implied that items such as school accreditation and teacher age did not play a role. Little is known about the impact of school and teacher characteristics on teaching quality. The previous school and teacher effectiveness research has focused more on identifying school and teacher factors related to student outcomes than on factors affecting teaching quality. Given these findings, however, it was decided not to use school and teacher characteristics in estimating the effect of the interventions on teaching quality in the multilevel analysis.

Next, all school and student characteristics were found to make differences to and have significant correlations with student achievement. However, not all characteristics could be retained as covariates in the multilevel analysis. At the two levels, only province, school score on English, father's education, and father's job were considered to have sufficiently significant correlation coefficients. However, although the characteristics teacher degree, teacher major, and teacher teaching experience at the teacher level had significant correlations with student achievement and made a significant difference to it, their coefficients were all below the acceptable level. In the current literature these three characteristics are often considered as teacher qualifications, while they have been found to yield rather mixed inconclusive results. Based on these findings, province, school score on English, and father's education were retained in the multilevel analysis. Because father's job was positively correlated with father's education ($r = 44$, $p = .00$), father's job was excluded due to its lower correlation coefficient. In addition, student gender was added. Despite its small correlation, this item was included because of the strong support from the literature for the effect of gender on student achievement.

To continue, the correlations of the school, teacher, and student variables in this study with student motivation were very small, while only some of them were significant. This result has been confirmed by previous research (e.g. Van Damme, Opdenakker, Van Landeghem, De Fraine, Pustjens, & Van de Gaer, 2006). The characteristics with the highest correlation coefficients at these three levels were school score on English, teacher degree, student gender, and father's education. Apart from their correlation, these characteristics also made a difference to student motivation. Considering the previous research, which has shown the smaller effect

of schools on student motivation and the role of prior achievement as reflected in the school score on the English national exam, these characteristics were retained in the multilevel modeling. Furthermore, since there was practically no research available into the effects of teacher characteristics on student motivation, it was considered interesting to see whether teacher degree would play a role in student motivation in the multilevel analysis. Therefore, this characteristic was also included.

Finally, as mentioned in chapter two, the literature strongly supports the role of gender in student motivation, whereas it has been inconclusive as regards the influence of SES. So although the correlation coefficient of gender was below the acceptable level, we also chose to include this characteristic in our measurement of the effects of the interventions on student motivation in the multilevel analysis. The following table summarizes the characteristics retained for the multilevel analysis for the different outcomes.

Table 4.10 *The Characteristics Retained for the Multilevel Analysis*

Characteristics	Outcome		
	Teaching quality	Student achievement	Student motivation
School			
Province	X	✓	X
school national exam	X	✓	✓
Teacher			
Teacher degree	X	X	✓
Student			
Gender	X	✓	✓
Father education	X	✓	X

CHAPTER 5

THE EFFECTS OF THE INTERVENTIONS ON TEACHING QUALITY AND STUDENT OUTCOME AND THE RELATIONSHIP BETWEEN THESE VARIABLES

5.1 Introduction

This chapter presents the results of the multilevel analysis of the effects of the interventions on the intermediary variable and the outcome variables. In addition, it addresses the relationship between the improvement patterns of these variables. The first analysis aimed at testing the effects of the interventions on teaching quality as the intermediary variable. Although it could be argued that, based on the findings in chapter 4 (see Table 4.1), this analysis was not necessary anymore, we did this analysis for sake of completeness, testing the hypothesis on the effects of the interventions at the appropriate levels, i.e. teaching quality nested in time and nested in teachers. Because we hypothesized that teachers in the experimental groups would improve their teaching quality more over time than teachers in the control group, we included the interaction between time and the interventions, next to establishing the main effects of these variables. The second analysis examined the effects of the interventions on student outcomes: student achievement and student motivation. Similar to the teaching quality data, the student motivation data were nested in time. So, also in this analysis the effects the interactions between time and the interventions were established. Finally, it was also tested whether or not teaching improvement explained the improvement of student outcomes. In this case teaching improvement was the difference between the teaching quality at measurement one and measurement three.

5.2 The results of the Multilevel Analysis

5.2.1 The effects of the interventions on teaching quality

The data on teaching quality were nested into two levels: time (starting point, middle point, and end point) and teachers. As indicated in chapter 4, the correlations between the school and teacher characteristics and teaching quality were not significant and made no difference to this variable. Therefore, they were not included as explanatory variables in the multilevel analysis. Five models were tested in addition to the empty model. To explore the pattern of improvement over time, the first model included time linear to examine whether the teachers in the three groups had shown significant improvements over the school year. Non-linearity was tested in Model 2, which included both time linear and time quadratic. In the third model, the effect of time was tested by allowing time to vary across the teachers in the

random part. The fourth model investigated the effects of the two interventions. Finally, the interaction effects between time and the two interventions were tested to examine if the teachers in the intervention groups had improved more than those in the control group.

The empty model (Model 0) as shown in Table 5.1 reports the variance components on both the teacher and the time levels. The results show that there was slightly more variance among the teachers at the teacher level (57%) than within the teachers at the time level (43%). In other words, the teaching quality differed from one teacher to another and from time to time. Furthermore, the variance among the teachers was somewhat larger than within the teachers. Figure 5.1 represents the variance among the teachers as well as that within the teachers across the three measurement points. With respect to the variance among the teachers, Figure 5.1 shows that their teaching quality started and ended at different points: some teachers started and ended their curves at high and low points. Concerning the variance within the individual teachers, the figure suggests that there seemed to be teachers whose quality linearly increased and decreased and teachers who experienced a curvilinear pattern of growth.

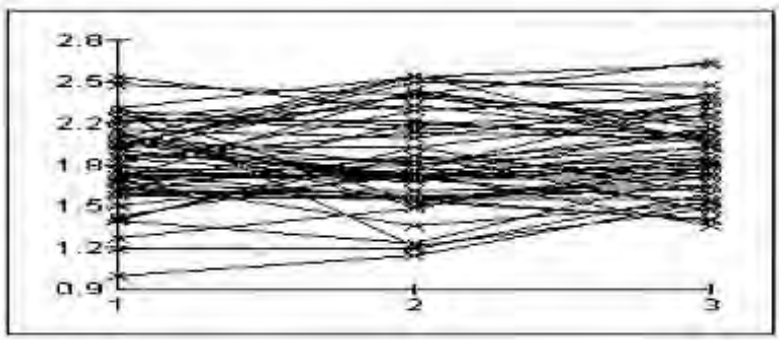


Figure 5.1 *The Variance of Teaching Quality within and among the Teachers across the Three Measurement Points (Empty Model)*

Table 5.1 *The Results of the Multilevel Analysis of the Effects of the Interventions on Teaching Quality*

	Model 0 (Empty Model)		Model 1 (Time)		Model 2 (Time Quadratic)		Model 3 (Time Fixed and Random)		Model 4 (Intervention)		Model 5 (Interaction Effect)	
	Coeff	SE	Coeff	SE	Coeff	SE	Coeff	SE	Coeff	SE	Coeff	SE
Fixed Part												
Intercept	1.902***	.039	1.828***	.057	1.868***	.134	1.828	.054	1.783***	.074	1.788***	.094
Time (linear)			.037	.021	-.011	.147	.037	.022	.037	.021	.035	.036
Time (quadratic)					.012	.036						
Intervention												
Intervention one									-.089	.087	.026	.135
Intervention two									.197**	.084	.086	.129
Interaction Effect												
Time X Intervention one											-.059	.052
Time X Intervention two											.057	.049
Random Part												
Teacher variance												
Intercept	.062	.015	.062	.015	.062	.015	.053	.017	.046	.015	.048	.015
Time							.003	.003	.000	.002	.000	.002
Time variance												
Intercept	.047	.007	.046	.006	.046	.006	.043	.007	.046	.007	.044	.006
Deviance (-2*loglikelihood)	49.713		46.634		46.525		45.499		35.467		30.349	
Decrease in deviance			3.079		.109		4.214		10.032***		5.142	
Variance explained by the model			.01		.00		.12		.04		.01	

** $p < .01$ (1-tailed), *** $p < .01$ (2-tailed)

As could already be expected from the descriptive analyses where a significant improvement took place only in experimental group 2 from measurement one to two, time as tested in the other models did not show significant effects. Time linear as included in Model 1 indicated that in general teachers neither improved nor declined significantly in their performance over time, at $p < .05$. Time linear explained 1% of the variance at the time level, although compared with the empty model, the model fit as shown by the decrease in deviance was not significant. Referring to the analyses of difference and effect size in chapter 4, the smaller variance at time level and the insignificant effect of time linear might be influenced by the fact that positive changes happened only in experimental group 2 from measurement one to two whereas stability took place in experimental group 1 and the control group over the three measurements.

Furthermore, the pattern of improvement over time was further investigated by including both time linear and time quadratic in Model 2, neither of which produced any effect. Because the model fit in model 1 was not significant, that of Model 2 was compared to the empty model, which neither appeared to be significant. The findings in Model 1 and Model 2 implied that neither the linear nor the polynomial function of time were appropriate to describe the changes of teaching quality over the three measurement across one school year.

In the next model, although the effect of time linear in the fixed part was not significant, following the hypothesis that time would have differential effects across teachers due to participating in the interventions, time linear was tested in the random part to see whether the effect differed per teacher. The results showed that the effect of time linear did not differ among the teachers. Based on Model 3, Figure 5.2 depicts that the teachers overall had the same pattern of change, which means that none of the individual teachers improved or declined more or less than the other. Because the model fit in Model 2 was not significant, it was compared to that of the empty model. Similar to the previous two models, also here the model fit was not significant. However, compared to the empty model (Model 0), this model explained 12% of the total variance. Given the large variance explained, this model could be considered as a better model than the empty model (Model 0) and was used as the reference of comparison with the next model.

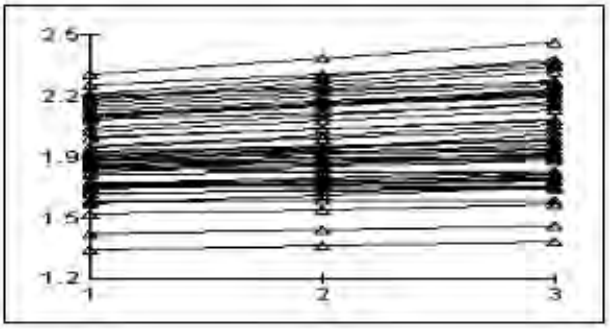


Figure 5.2 *The Growth of the Teaching quality of the Individual Teachers Over Time (Model 3)*

The fourth model introduced the two interventions as the explanatory variables while controlling for time linear because these data included time as a level one variable. Time quadratic was excluded due to its insignificant result as shown by Model 2. Time linear was also included in the random part because the inclusion of this variable as tested in Model 3 explained a large amount of variance (12%) although the model fit did not show that it was a better model compared to the empty model. Similar to Model 1, time linear in this fourth model was not significant and the teachers did not differ in their growth over time. As regards the interventions, the results revealed that the effect of the first intervention was not significant. As shown in Table 5.1, there was no difference between experimental group 1 and the control group. However, the coefficient of the second intervention was positive and significant. The teaching quality of the teachers in experimental group 2 was .199 higher than in the control group. Because Model 3 was considered a better model than the empty model due to its large explanation to the variance, we compared the deviance in this model to that in Model 3, and found a large decrease of 10.032 (*df*= 2), which suggested that the model fitted the data well. The interventions themselves added 4% to Model 3 in the explaining of variance, and hence the model explained 16% of the total variance. In line with this finding, decreases of variance on both level one and level two were noticed.

In order to examine whether throughout the school year the teachers in the experimental groups had progressed more than the teachers in the control group,

the final model included the interaction effect between time and the two interventions. One could argue that this analysis was not necessary given that there was no fixed effect of time. However, for the sake of completeness, this analysis was performed to test the hypothesized interaction effects between time and the interventions. The results as described in Table 5.1 indicate that both interaction effects were not significant, implying that across time the two experimental groups had not improved their teaching quality more than the control group. In addition, compared to Model 4, the model fit was not significantly better (5.142, *df* = 2).

In summary, the results show a variance in teaching quality across time within individual teachers, which suggests that teaching quality was not a stable condition within individual teachers. There was also variance in teaching quality among the teachers, which was somewhat larger than the variance within individual teachers. However, based on Model 4 as the most representative one, there was overall no significant increase in teaching quality over time. These findings were likely due to the fact that progression was observed only in experimental group two (only from measurement one to two), whereas stability was noticed in both experimental group 1 and the control group over the three measurements (see Table 4.1, chapter 4). In addition, the effect of time did not differ from one teacher to another. On average, the analysis revealed that across the three measurement points the level of teaching quality was the highest among the teachers in the second intervention group. Nevertheless, the most important concern in this study was the improvement of teaching quality over time of the intervention groups, which was tested in Model 5. The results show that in both experimental groups the teachers did not improve more over time than the teachers in the control group, which is why we had to reject the hypothesis concerning the effect of the interventions on teaching quality. These results confirm those which already were found in the descriptive findings (see Table 4.1).

5.2.2 The effects of the interventions on student outcome

The cognitive aspect of the student achievement in reading comprehension was measured twice (pretest and posttest). Two perspectives were used: an analysis of covariance (posttest) while controlling for prior achievement (pretest) and an analysis of learning gain. This approach was intended to provide a more

comprehensive description of the intervention effects on student achievement. Experimental studies, in which the subjects are randomly assigned into groups, regardless of the bias in the initial measurement, require an analysis of covariance. A learning gain analysis is important when change is considered, which also applied to this study. In view of the selection of the characteristics at the different levels which could be retained in the multilevel analysis as described in chapter 4, different explanatory variables were introduced for estimating the effects of the interventions on both outcomes. The difference was the exclusion of the pretest as the covariate in estimating learning gain.

As opposed to student achievement, student motivation as a non-cognitive outcome in this study was measured three times and nested on three levels: time (starting point, middle point, and end point), student and school. Therefore, multilevel growth modeling was employed to obtain an understanding of the changes in student motivation over time and to investigate if the experimental groups had produced more favorable results than the control group. In estimating this non-cognitive outcome variable, student and school as well as teacher characteristics were included as explanatory variables. The following subsections present the student achievement and the student motivation analyses.

5.2.2.1 The effects of the intervention on student achievement

The covariance analysis

For analyzing the students' posttest scores, two models - one including student covariates and the other school level covariates - were used to examine the unique variance on these levels. The third model tested the effects of both the student and the school level covariates. Next, Model 4 measured the effects of the interventions while controlling for the covariates tested in Model 3. Then, the first two models were compared with the empty model to estimate the model fit and the variance explained by each of them. This procedure was carried out to determine the unique variance explained by each level. The third model combined both student and school characteristics, and was also compared to the empty model. Finally, Model 4 – which tested the effects of the interventions - was compared to Model 3 to see how much more variance the two interventions could explain.

Concerning the variance component, the random part in the empty model (Model 0) showed that at the student level it was 67% and at the school level 33%. Hence, based on their posttest score the students differed twice as much in reading comprehension as the schools did. As expected, the inclusion of explanatory variables on both the student and the school levels in the next models led to a decrease in variance at both levels, which suggests that the explanatory variables played a role in explaining these variances.

The first model introduced student characteristics as explanatory variables. They included prior achievement (pretest), gender, and father's education. In this study, father's education contained four categories, from primary school to university education. Here the regular contrast was used, so the lowest category was the reference of the comparison. The analysis showed that prior achievement and gender had significant effects on the students' reading achievements. Students who achieved a high score in the pretest also achieved a high score in the posttest. In addition, female students achieved significantly higher than the male students. Surprisingly, SES as represented by father's education, was not significant at $p < .01$, as set for the student data in this study. Furthermore, the coefficient of students whose father had graduated from the university was not significant, although it was in the expected direction. Compared to the empty model (Model 0), the decrease in deviance was significant (1452.751, $df = 5$) while the above student characteristics explained 6% of the total variance on both the student and the school levels.

The second model included the effects of the school covariates: province and school prior achievement (the school score on the English national exam). Province and a high school prior achievement (8.1 – 9) were found to have significant effects on the posttest scores, province having a negative and school prior achievement a positive impact. With respect to province, this finding meant that students in DKI Jakarta, the capital city of Indonesia, scored significantly better than those in Banten. As regards school prior achievement, students in schools with a high score on the English national exam achieved significantly better in reading than students in schools with a low score on the English national exam. Compared to the empty model (Model 0), the model fit showed that this model was a better one, showing a decrease in deviance of as much as 383.587 ($df = 3$). As Table 5.2 indicates, the

variance decrease was quite large at the school level and small at the student level, while the model explained 11% to the total variance.

In the next model, all covariates at both the student and the school levels were combined. We saw that the significant covariates tested in the previous two models remained significant, while the model fit showed that it was a better model than the empty model (Model 0). The student and school level covariates together explained 16% of the total variance, showing considerably more decrease in variance at the school level than at the student level.

Next, the above model was added to the two interventions (Model 4). The results show that the second intervention was associated with better student achievement rates in reading, which were significant at $p < .01$ (1-tailed). Although the students in the first intervention had also improved their achievements, this result was only significant at $p < .05$ (1-tailed). The model fit as displayed in the decrease in deviance (9.009, $df = 2$) showed that Model 4 was a better Model than 3. The two interventions explained 4% more variance, which made model 4 explain 20% of the total variance.

In summary, the results tell us that when controlling for various characteristics at the student and the school levels, the students' posttest scores in experimental group 2 was significantly better than that in the other two groups. The coefficient of the first intervention went in the expected direction but was not significant. The characteristics found to be significant predictors of the students' posttest scores included student prior achievement, gender, province, and school prior achievement.

Table 5.2 *The Results of the Multilevel Covariance Analysis*

	Model 0 (Empty Model)		Model 1 (Student)		Model 2 (School)		Model 3 (Student + School)		Model 4 (Intervention)	
	Coeff	SE	Coeff	SE	Coeff	SE	Coeff	SE	Coeff	SE
Fixed Part										
Intercept	9.468***	.270	7.872***	.401	9.584***	.491	8.253***	.563	7.389***	.609
Student level										
Pretest			.149***	.028			.146***	.028	.146***	.028
Gender (female)			.442***	.153			.395***	.155	.397***	.155
Father education (JHS)			-.186	.216			-.211	.216	-.223	.216
Father education(SHS)			-.055	.209			-.114	.212	-.136	.212
Father education (univ)			.684	.333			.560	.339	.510	.339
School level										
Province (Banten)					-1.976***	.484	-1.955***	.468	-1.648***	.444
School nat. exam (med.)					.907	.551	.691	.539	.515	.503
Schoolnat. exam (high)					2.791***	.924	2.462***	.892	1.877***	.862
Intervention										
Intervention one									.904	.545
Intervention two									1.679**	.537
Random Part										
School level variance										
Intercept	3.875	.784	3.457	.715	2.703	.574	2.423	.531	2.003	.450
Student level variance										
Intercept	7.773	.275	7.476	.291	7.639	.275	7.317	.292	7.316	.292
Deviance (-2*loglikelihood)	8268.439		6815.688		7884.852		6455.230		6446.221	
Decrease in deviance			1452.751***		383.587***		1813.209***		9.009***	
Variance explained			0.06		0.11		0.16		0.04	

**$p < .01$ (1-tailed), *** $p < .01$ (2-tailed)

The learning gain analysis

Learning gain was the difference between the posttest and the pretest scores. As previously explained, this analysis was carried out to investigate the possible pretest-posttest changes in student achievement. Here we used the same covariates and procedures as employed in the former analysis, of course with exception of pretest score at the student level.

The intercept in the fixed part of the empty model (Model 0) showed that in general there was no learning gain. Nevertheless, the random part displayed a significant variance at both the school and the student levels. At the student level (84%) this variance was around four times larger than at the school level (16%). So in the next models, both student and school characteristics were included to test whether they contributed in explaining this variance.

The first model introduced student characteristics and demonstrated that none of them were significant. The model fit improvement as shown in the decrease in deviance (1546.092, *df* = 4) was significant, which indicated that it was a better model. However, the variance at school level increased to 2.752 from 2.445 in the empty model (Model 0), thereby explaining the variance negatively (-0.01). The second model included the effect of school characteristics, with province as the only significant covariate. The coefficient was negative, which suggested that the learning gain of students in the Banten province was significantly lower. This result was similar to that of the covariance analysis, so students in DKI Jakarta had a higher posttest score and a higher learning gain. Compared to the empty model, the decrease in deviance was significant (383.219, *df* = 3), while the model explained 3% to the total variance.

Next, all covariates at both the student and the school levels were combined in Model 3, which yielded the same results as produced by the previous two models except for province, which had a significant negative effect. Compared to the empty model (Model 0), the decrease in deviance (1910.206, *df* = 7) suggested that it was a better model. Since Model 1 had shown no effects of student characteristics, it was not surprising that the combination of student and school covariates did not add any additional explanation. So also Model 3 explained 3% to the total variance, which referred to the school covariates as examined in Model 2.

The effects of the two interventions were tested in Model 4, while controlling for all covariates tested in Model 3. Also here the impact of the second intervention was confirmed; the same outcome as calculated in the covariance analysis. As regards the covariates, only province had a significant negative effect, as also indicated by Model 3. However, compared to Model 3 the decrease in deviance was not significant. Nevertheless, considering the fact that student characteristics did not explain any variance and school characteristics explained only 3%, this model was important because it explained 2% more of the total variance, which amounted to a total of 5%.

In brief, although in general no learning gain was observed, there was variance at the student and the school levels. Similar to the covariance analysis, the learning gain analysis confirmed the effect of the second intervention while controlling for various background characteristics. As regards these background characteristics, only province played a role in the students' learning gain. This coefficient was negative, which was also the case in the covariance analysis. Therefore, it could be concluded that both the posttest and the learning gain scores of the students in the Banten province were lower than those of the students in DKI Jakarta.

Table 5.3 *The Results of the Multilevel Learning Gain Analysis*

	Model 0 (Empty Model)		Model 1 (Student)		Model 2 (School)		Model 3 (Student + School)		Model 4 (Intervention)	
	Coeff	SE	Coeff	SE	Coeff	SE	Coeff	SE	Coeff	SE
Fixed Part										
Intercept	-.062	.227	.137	.328	.619	.445	1.092	.530	.548	.597
Student level										
Gender (female)			.172	.198			.134	.203	.133	.203
Father education (JHS)			-.395	.280			-.415	.284	-.438	.284
Father education (SHS)			-.431	.270			-.513	.277	-.556	.277
Father education (Univ)			-.156	.429			-.370	.441	-.458	.442
School level										
Province (Banten)					-1.395***	.435	-1.590***	.464	-1.389***	.455
School nat. exam (medium)				-.094	.497	-.232	.534	-.331	.517	
School nat. exam (high)					.777	.832	.734	.880	.398	.882
Intervention										
Intervention one									.412	.557
Intervention two									1.230**	.553
Random Part										
School level variance										
Intercept	2.445	0.546	2.752	0.625	1.942	0.462	2.118	0.519	1.883	0.473
Student level variance										
Intercept	12.874	0.455	12.728	0.495	12.858	0.464	12.687	0.506	12.685	0.505
Deviance (-2*loglikelihood)	9057.317		7511.225		8674.098		7147.111		7142.012	
Decrease in deviance			1546.092***		383.219***		1910.206***		5.099	
Variance explained			-0.01		0.03		0.03		0.02	

** $p < .01$ (1-tailed), *** $p < .01$ (2-tailed)

5.2.2.2 The effects of the interventions on student motivation

In estimating the effects of the interventions on student motivation, several models were tested. After the empty model, different functions of time were investigated to explore the patterns of changes in student motivation over time. Firstly, time linear was introduced to examine whether the students in the three groups had improved their motivation over a school year interval. Secondly, the non-linear function of time was tested in another model by including time quadratic. Also related to time, the next model examined whether the effect of time linear differed across the schools by including this component in the random part.

Another three models tested the contribution of the background characteristics at the student, teacher, and school levels separately in order to estimate the unique variance explained by each level. However, only Model 4, which combined all of these covariates, has been presented due to a lack of space. After that, the effects of the interventions were tested in Model 5 while controlling for various characteristics. Furthermore, the interaction effects between time and the interventions were tested to investigate whether the students' motivation in the experimental groups had progressed more than that in the control group.

As Table 5.4 illustrates, in this analysis the largest variance shown in the random part of the empty model (Model 0) occurred at the lowest level: time (73%). The variance at the student level was 20% while the smallest amount was observed at the highest level, namely school (7%). Hence, student motivation differed from time to time within individual student, from one student to another and from one school to another. Given the large variance at the time level, student motivation was an unstable characteristic within individual students. However, as investigated in the first two models, the pattern of motivation over time in general showed both non-significant decreasing pattern of linear time effect and non-significant quadratic function of time. So these models did not explain any variance and neither did they prove to be better models than the empty model.

Table 5.4 *The Results of the Multilevel Analysis of Student Motivation*

	Model 0 (Empty Model)		Model 1 (Time Linear)		Model 2 (Time Quadratic)		Model 3 (Time Fixed and Random)		Model 4 (Student, Teacher, School)	
	Coeff	SE	Coeff	SE	Coeff	SE	Coeff	SE	Coeff	SE
Fixed Part										
Intercept	2.939***	.019	2.960***	.025	2.925***	.053	2.963***	.026	2.772***	.520
Time (linear)			-.011	.008	.032	.057	-.013	.010	-.012	.011
Time (quadratic)					-.011	.014				
Student level										
Gender (female)									.062***	.018
Teacher level										
Teacher degree (bachelor)									.093	.057
Teacher degree (master)									.219***	.086
School level										
School nat. exam (medium)								.070	.052	
School nat. exam (high)									.236***	.080
Random Part										
School level variance										
Intercept	.013	.003	.013	.003	.013	.003	.014	.005	.008	.004
Time							.002	.001	.002	.001
Student level variance										
Intercept	.041	.004	.041	.004	.041	.004	.042	.004	.041	.004
Time level variance										
Intercept	.148	.004	.148	.004	.148	.004	.145	.004	.145	.004
Deviance (-2*loglikelihood)	3893.954		3892.257		3891.689		3881.384		3762.435	
Decrease in deviance			1.697		2.265		12.570***		131.519***	
Variance explained			0,00		0.00		0,00		0,04	

*** $p < 0.1$ (2-tailed)

Table 5.4 *The Results of the Multilevel Analysis of Student Motivation*

	Model 5 (Intervention)		Model 6 (Interaction Effect)	
	Coeff	SE	Coeff	SE
Fixed Part				
Intercept	2.716***	.055	2.665***	.057
Time	-.013	.011	.048*	.019
Student level				
Gender (female)	.061***	.018	.060***	.018
Teacher level				
Teacher degree (bachelor)	.111*	.056	.101	.053
Teacher degree (master)	.236***	.082	.218***	.079
School level				
School nat. exam (medium)	.061	.051	.053	.049
School nat. exam (high)	.210*	.081	.192*	.077
Intervention				
Intervention one	.034	.050	.124	.058
Intervention two	.108**	.047	.212**	.056
Interaction effect				
Time X intervention one			-.079**	.025
Time X intervention two			-.092**	.025
Random Part				
School level variance				
Intercept	.006	.003	.006	.003
Time	.002	.001	.002	.001
Student level variance				
Intercept	.041	.004	.041	.004
Time level variance				
Intercept	.144	.004	.144	.004
Deviance (-2*loglikelihood)	3757.161		3743.171	
Decrease in deviance	5.274		19.264***	
Variance explained	.02		.02	

* $p < .05$ (2-tailed), ** $p < .01$ (1-tailed), *** $p < 0.01$ (2-tailed)

In Model 3, time linear was included in both the fixed and the random part to see whether the effect of time differed across the schools. Whereas in the fixed part time linear showed a non-significant negative effect, in the random part it had a significant effect, which suggests that its effect differed per school. Although the model did not explain any variance, it was better than the empty model (Model 0), given the decrease in deviance (12.570, *df* = 2).

The next model presented in Table 5.4 examined the combined effects of the student, teacher, and school level covariates. Before that, separate analyses had been conducted to measure the effects of each level, the results of which were similar to those in Model 4. At the student level, student gender was found to have a significant effect, which means that the motivation of the female students was significantly higher than that of the male students. At the teacher level, teacher education at the Master level was found to have a significant effect on motivation. This finding suggests that students whose teachers had finished a Master degree were more motivated than those whose teachers had completed only diploma degree. In Model 4, the teacher bachelor degree was not associated with a higher level of student motivation. However, in the previous analysis, in which only the effect of teacher characteristics was tested, both bachelor and master degrees were found to contribute to better student motivation outcomes. At the school level, the high school's prior achievement was positively related to high student motivation. With respect to time, the result was the same as that in Model 3, while overall there was no time linear effect in the fixed part. However, as shown in the random part, there was variance in the time linear effect across the different schools.

As regards the model fit, the decrease in deviance (131,519, *df* = 7) made Model 4 more suitable than the empty model (Model 0). It was able to explain 4% of the variance in student motivation, which was caused by teacher and school level covariates. When testing the effect of covariates at the different levels in the three separate models, only the student covariate (gender) was found not to explain any variance.

Next, the effects of the two interventions were investigated in Model 5. As we see in Table 5.4, similar to the previous analysis of student achievement, the students in the second intervention group had on average a higher level of motivation across

the three measurement points. In addition, teacher bachelor degree was also found to be associated with a higher student motivation rate. This model added 2% to the explained variance, but the model fit (5.275, *df* = 2) indicated that it was not a better model than Model 4. However, the interaction effects between time and the interventions as tested in Model 6 indicated that both experimental groups experienced a significant decline over time. This finding means that the two interventions had no effect on student motivation. Furthermore, this model yielded an interesting result: time linear became significant at $p < .05$ (2-tailed). Because in terms of model fit, Model 5 was not significant, Model 6 was compared to Model 4, in which all covariates had been combined. Here the model fit, as shown in the decrease of deviance (19.264, *df* = 5), was significant. The very small decrease in the variance components at the time and teacher levels suggested a very small influence of the interventions, the interactions, and the teaching improvement. They explained another 2% of the total variance.

5.2.3 The mediating role of the improvement of teaching quality in the effects of the interventions on student outcome

The multilevel analysis has shown that the teaching quality in the second intervention group was higher than in the first intervention group and the control group. Nevertheless, none of the intervention groups improved more over time compared to the control group. In terms of student outcome, the covariance and the learning gain analyses demonstrated that the second intervention had a significant effect on student achievement. However, as regards student motivation no effects were observed, neither in the first nor in the second intervention. Therefore, given these findings, teaching quality improvement had strictly speaking no mediating role. The reason why we performed this analysis was for the sake of completeness. Appendix 5.1 provides the results.

For each of the analyses of student achievement and student motivation, three models were developed. The analysis is provided in the appendix, in which these three models have been labeled as Model 5, 6, and 7 for the covariance and learning analyses, and as Model 7, 8, and 9 for student motivation. Firstly, we only included the covariates and teaching improvement in the model (Model 5 for covariance and the learning gain analyses and Model 7 for motivation) to examine the unique

explanation of teaching improvement of the variance in the covariance (posttest), learning gain, and motivation analyses. Therefore, each model was compared to the previous model which tested the combination of covariates at the different levels (Model 3 for the covariance and learning gain analyses and Model 4 for student motivation). The results revealed that teaching improvement was neither associated with student achievement (covariance and learning gain) nor with student motivation. The model fit for these outcome variables was not significant and did not explain any additional variance. The exception was that in student motivation, which added 1% in explaining the variance.

Secondly, the interventions were included. For the covariance and learning gain analyses, each model was compared to Model 4, which had tested the effects of the interventions while controlling for background characteristics. As regards student motivation, Model 8 was compared to Model 6, which not only tested the interventions but also the interaction between the interventions and time. The results of these three analyses were the same. Teaching improvement did not mediate the effects of the interventions on student outcome while the model fits of the analyses were also not good. Finally, the interaction effect between the interventions and teaching improvement on student outcomes was examined. The results indicated that the effects of the second intervention, which as we knew were significant as shown by both the covariance and the learning gain analyses, had not been mediated by the improvement of teaching quality in this group.

5.3 Summary of the research findings

This chapter has attempted to answer the three main research questions posed in this study through multilevel modeling analysis. In summary, the findings showed that only the second intervention had significant effects, which were merely visible in student achievement, as evidenced by the covariance and learning gain analyses. The second intervention was also found to have a negative interaction effect with time on student motivation.

With regard to teaching quality, in general no improvement was observed over time, which could due to improvement found only in experimental group 2 in the first half of the teaching improvement program and stability in the other two groups. Furthermore, neither did the teachers' teaching quality differ in terms of progression

over time. The teachers in experimental group 2 were on average found to have significantly better teaching quality but their teaching quality over time did not improve more compared to those in the other two groups. Therefore, it was concluded that none of the interventions, including the second one, had an effect on teaching quality. From the descriptive data in chapter 4, we could see that there was a significant improvement in experimental group 2 from measurement one to two, but this group did not improve further from measurement two to three. Moreover, teaching improvement was not associated with student outcome. In other words, although the students in experimental group 2 had better posttest scores and more learning gain, this outcome was not explained by teaching improvement.

Furthermore, although the motivation of the students in the second intervention group was found to be significantly higher, it declined over time. So the interventions also did not have an effect on student motivation. Therefore, only one hypothesis in this study was confirmed: the second intervention had a positive effect on student achievement, as demonstrated by both the covariance and the learning gain analyses.

CHAPTER 6

SUMMARY, CONCLUSIONS, AND DISCUSSION

6.1 Introduction: research background and theoretical background

In 2005 the Indonesian government formally established the education standards as a national strategy to improve the quality of education. One important objective was to enhance student outcome. Eight education standards were formulated, two of which could be considered as the principal achievement directives, the standards of performance and the standards of content. In the Indonesian context, the standards of content are a more detailed explanation of the standards of performance.

Educational standards have been claimed to promote equity and excellence because their main focus is to provide all students, regardless of their background, with the opportunity to achieve the same minimum educational objectives (e.g. Sandoltz, Ogawa, & Scribner, 2004; Buttram & Waters, 1997; McClure, 2005; Wurtz et al. 1993). Without educational standards, the teaching approaches conducted by the teachers within a particular subject domain or grade may largely differ, depending on their individual knowledge and views, as well as their personally selected textbooks and/or other teaching materials. Individual teachers may also use different expectation criteria for different students. This situation strongly increases the degree of differentiation among the curricula, the methods of assessment and the student performance rates. When using educational standards, however, teachers will have to apply the same expectation criteria to all students, thereby encouraging all of them to acquire the same minimum competences.

However, as in other countries, not much has been done yet to examine the effectiveness of these standards. Dowson et al. (2000), for example, concluded that there is hardly any scientific knowledge of whether the claims as regards equity and excellence are true. On the other hand, some studies have shown that the standards movement has only affected certain groups, and that particularly the disadvantaged students have not benefited from the policies in this field (e.g. Lauer et al. 2005; National Research Council, 2001; Thompson, 2009; Zuzovsky & Libman, 2006). Two underlying problems have been identified: the broad and vague formulations of the standards (Gandall, 1996; Finn, Petrili, & Vanourek, 1998) and the lack of focus on the teachers (Chamber & Dean, 2000; Mathison & Freeman, 2008). Also in Indonesia the content descriptions of the standards have been

problematic, while teachers have had serious difficulties with translating them properly into their teaching and learning activities.

Although at the present moment there is not much knowledge available from standards-based education of how to improve education and solve the issues described above, it is clear that educational enhancement is crucial, both to tackle the low quality levels and to maintain the improvements once they have been realized. It is therefore important for policy makers to help schools find strategies that could facilitate this educational improvement. In this context, the Educational Effectiveness Research (EER) has offered both theory-driven and evidence-based information on what could be prioritized in educational improvement. It has been considered as a body of research harboring different areas of investigation, such as teacher behavior, the curriculum, grouping procedures, school organization, and educational policy (Creemers & Kyriakides, 2008). A considerable number of studies in these fields have confirmed the superiority of the classroom or teacher level (e.g. Creemers, 1994, Darling-Hammond, 1997; Harris & Muijs, 2005, Hill & Rowe, 1996; Luyten & Snijders, 1996; Marzano, 2007; Van der Werf, et al. 2000). Thereby, the importance of prioritizing improvement measures at this particular level as a strategy to advance education has been emphasized. Moreover, this approach simultaneously addresses the second issue of standards-based education: the role of the teachers.

Based on the summary of the literature review in this study, two important propositions could be formulated. The first was the importance of the education standards, especially in view of formulating shared expectations about what students should learn and accomplish. However, harmonization of the learning expectations requires standards which are specific, clear, and concrete. Only then can the teachers adequately translate these standards into effective learning methods and help the students achieve their educational targets. The second issue concerned the importance of assisting the teachers in improving their teaching quality, especially with respect to the factors related to student outcome as indicated by EER. It was argued in this study that the combination of the standards-based education and a teacher improvement program might yield better results than the standards-based education approach on its own.

Based on these two premises, two interventions were developed and compared in terms of their effectiveness in improving teaching quality and student outcome. Prior to the interventions, two supporting documents were developed. The first was referred to as the elaborated standards document, which offered supplemental explanations of the text in the original standards document. It was an attempt to make the government's standards more concrete and unambiguous. Here the elaborated standards specifically referred to the standards of content, since in the Indonesian context these are a more detailed explanation of the standards of performance. The second document included the characteristics of effective teaching as identified by EER. In the first intervention only the elaborated standards document was provided to the sample group. In the second intervention the participants could make use of both the elaborated standards document and the teacher improvement information. Furthermore, they attended a teacher improvement program.

For designing the teacher improvement program in the second intervention, the dynamic model of educational effectiveness (Creemers & Kyriakides, 2008) was used, in which the focus is especially on elements at the classroom level. Although this model includes different levels, namely national/regional, school, classroom/teacher and student, it emphasizes the classroom level while uses the higher levels to provide the necessary conditions for the effectiveness of the classroom level. Findings of previous studies, especially conducted in Cyprus, have supported the validity of this model at both the school and the classroom levels (Antoniou, Demetriou & Kyriakides, 2006; Antoniou, 2009).

Eight factors have been distinguished at the classroom level, all of which refer to the observable teacher instructional roles, considered by previous teacher effectiveness studies to be related to student outcome. They are orientation, structuring, modeling, application, questioning, creation of the classroom as a learning environment (CLE), time management, and assessment. These factors cover various teaching approaches, such as constructivism and direct instruction/ mastery learning. An experimental study based on these classroom factors showed an increase in teaching quality and student achievement (Antoniou, 2009). Although its design differed from that of Antoniou (ibid), the teacher improvement program in this study also included all classroom factors of the dynamic model.

In testing the effectiveness of the two interventions, a control group was formed for the purpose of comparison. Three main research questions were posed in this study: 1) What are the effects of an elaborated standards intervention and those of an intervention in which the elaborated standards are combined with a teacher improvement program on improving teacher's teaching quality? 2) What are the effects of and elaborated standards intervention and those of an intervention in which the elaborated standards are combined with a teacher improvement program on improving student outcome? 3) To what degree can the effects of the interventions on student outcome be explained by an improvement of the teaching quality?

Based on these questions, three hypotheses were formulated. The first was the assumption that if teachers have clear ideas about the targets which students are required to accomplish, the improvement of teaching quality and student outcome will be higher than when the original standards document provided by the Indonesian government is used. The second hypothesis was that if teachers are supported by a combination of an improvement program and a clear and specific standards document, the improvement of teaching quality and student outcome will be higher than when they are only provided with a clear and specific standards document. The final hypothesis was that the effects of the interventions on student outcome can be explained by the improvement in teaching quality.

6.2 Research methodology

For our research we used a pretest-posttest randomized experimental design, which incorporated a longitudinal approach in which there were three points of measurement within one school year. The study focused on the subject English (reading comprehension), in madrasah Tsanawiyah[2]. Madrasah is a type of Junior Secondary School (JSS) under the Indonesian ministry of religious affairs, located in the two provinces DKI Jakarta and Banten. This type of school was selected because of its lower quality of education and hence its urgent need for assistance in the realization of improvement. JSS is a three-year schooling trajectory after six years of primary school. Only the second year students were included in the study,

[2] Indonesia has a dual schooling system. One component is the general school supervised by the Ministry of Education and the other is the madrasah, supervised by the Ministry of Religious Affairs. They have the same curricula and levels, but madrasah offers additional religious subjects.

because the first year students had just started their English lessons and the third year students had to prepare for the national exam.

The participants voluntarily participated and were randomly assigned to two experimental groups and one control group. In the first experimental group the first approach was used, and in the second group the second method. The control group worked with the standards documents from the government. Prior to the interventions two supporting documents were developed. The first was named “the elaborated standards of English”, which was an attempt to clarify the government standards (of content) by providing specific definitions of the reading competencies. The second was titled “becoming an effective teacher of English”, which contained both research-based and practical strategies to improve teaching quality by referring to both the reading skills in the elaborated standards and the classroom factors of the dynamic model as presented by EER.

So the first experimental group used the elaborated standards document and attended a one-day workshop to discuss its content. In line with the standards-based education directives, the teachers in this group were advised to search themselves for strategies to enable them to facilitate their students in achieving the goals as targeted by the standards. The second experimental group was provided with both the elaborated standards and the document on becoming an effective teacher of English. In addition to the one-day workshop on the elaborated standards, they also attended an instruction on effective teaching and six half-day monthly meetings as part of the teacher improvement program as designed in this study. The contents of the workshop and the monthly meetings meant for the second intervention group were based on the elaborated standards document and the classroom factors of the dynamic model.

We measured an intermediary variable at the teacher level (teaching quality) and outcome variables at the student level (student achievement and student motivation). In measuring teaching quality, the classroom level of the dynamic model was used as a conceptual framework in which two instruments were used: an observation instrument and a student questionnaire, both developed based on the aforementioned studies in Cyprus (Creemers & Kyriakides, 2008). Two pilot studies were conducted to adjust these instruments to the Indonesian context and the

specific subject dealt with in this study. The classroom observation was carried out by independent observers who had previously been trained and had attained a good inter-rater reliability. Each observation was done by one observer. Student achievement was measured through a reading comprehension test. We used the reading comprehension test level A2 of the Central Institute for the Development of Tests (CITO) as a starting point, but modified and reduced it to make it suitable for the reading skills as defined in accordance with the elaborated standards and the Indonesian context. Finally, student motivation was measured via a student questionnaire originally developed by Hermans (1983) and later modified by Kuyper, Van der Werf & Lubbers (2000).

The study was conducted during one school year, from mid July 2010 until the end of May 2011, during which data were gathered three times. In measuring teaching quality, three classroom observations were conducted by independent observers, while the questionnaire on the students' perception of the teaching quality was distributed two times, namely during the second and the third measurements. Student achievement was tested by administering a pretest and a posttest at the beginning and at the end of the study, respectively. Finally, student motivation was measured three times.

The sample population consisted of in total 57 schools, including 59 teachers (M = 44%, F = 56) and 2,431 students (M = 48,5%, F = 51,5%). Each measurement contained a different number of participants: 1,660 students (54 schools, 56 classes) participated in both the pretest and the posttest, while 1,133 students (50 schools, 52 classes) were included in the three measurements of student motivation. The student questionnaire on teaching quality was distributed in measurement two and three among 1,264 students (43 schools, 45 classes) who completed both sessions. With respect to the second intervention, the teacher improvement program was joined by only half of the participants due to various kinds of personal and practical problems.

In addressing the research questions of this study, both descriptive and multilevel analyses were performed. The descriptive part concerned the differences among the three groups as regards the intermediary variable and the outcome variables. The multilevel analysis was carried out to answer the research questions.

6.3 Summary of the research findings in relation to the research questions

The summary of the research findings presented in this section will focus on the results of the multilevel analysis, which was the predominant measurement performed to investigate the research questions. We have to add, however, that the descriptive findings have certainly also been considered to be supporting the results. The summary is based on the intermediary variable and the outcome variables measured in this study. In line with the research questions, teaching quality will be first presented, followed by student achievement and student motivation. The third question about the relationship between teaching improvement and student outcome will be explained when discussing student achievement and student motivation.

6.3.1 Teaching quality

The data on teaching quality were nested within 'time' and 'teacher', thereby employing a multilevel growth modeling approach. The multilevel analysis reported that from one measurement to the next the teaching quality within the individual teachers varied almost as much as that from one teacher to another. Time did not have a significant effect. Furthermore, the multilevel analysis revealed that neither did the teachers across the different schools vary in their growth of teaching quality over time.

With respect to the interventions, the participants in experimental group 2 were on average found to have better teaching quality. However, as indicated by the interaction effects, there were no differences in quality improvement between the experimental groups and the control group. Hence, despite the fact that the teachers in intervention group 2 were better, teaching quality in this group did not improve more compared to that of the teachers in the other two groups. Therefore, in terms of teaching quality, both hypotheses in this study had to be rejected: the teaching improvement in experimental group 1 was not larger than that in the control group and in experimental group 2 it was not larger than in experimental group 1 or the control group. Nevertheless, the descriptive findings in chapter 4 indicate that there was a significant improvement in teaching quality in experimental group 2 from measurement one to two, which is important to note here. We will come back to this later in this chapter.

6.3.2 Student achievement

In investigating student achievement, a covariance and a learning gain analysis were performed to facilitate a more comprehensive examination of the intervention effects. In the multilevel analysis the background characteristics that had previously been found to be associated with student achievement, were included as covariates to calculate the intervention effects more precisely. These background characteristics were identified on two levels in the correlation analysis as part of the descriptive analysis. They were gender and SES at the student level, and province and school prior achievement at the school level, both of which were associated with student achievement. None of the teacher characteristics were found to be associated with student achievement.

On both the student and the school levels variance occurred in the covariance and the learning gain analyses. In both analyses this variance was much larger at the student level than at the school level. In these analyses the same covariates were used, except that student prior achievement was excluded from the learning gain analysis. In the covariance analysis (as part of the multilevel analysis) significant effects were found of student prior achievement, gender, province, and (high) school prior achievement. University as the father's education (representing SES) was in the expected direction but not significant. On learning gain, none of the student characteristics had a significant effect. Province was the only significant covariate for learning gain, which had a (negative) effect according to the covariance analysis: the students in Banten scored significantly lower on the posttest and on learning gain than the students in DKI Jakarta.

While controlling for the above background characteristics, the second intervention was found to have a significant effect on student achievement, both in the covariance and the learning gain analyses. The regression coefficient of the first intervention was in the expected direction but not significant. Hence, the hypothesis that the elaborated standards combined with a teacher improvement program would result in better student achievement scores than the strategies used by the first intervention group and the control group was confirmed. However, no relationship was observed between teaching improvement and the posttest scores nor learning gain.

6.3.3 Student motivation

Like teaching quality, student motivation was analyzed using multilevel growth modeling, because – similarly to those on the student and the school levels – the data were nested in time. The results reported that student motivation varied within the individual students, among the students, and among the schools. The variance within the individual student groups was much higher than that among the students and among the schools. This finding suggests that motivation was an unstable characteristic within the individual students. However, on average it neither progressed nor declined over the school year. Nevertheless, the effect of time (linear) differed across the schools, suggesting that there were schools where student motivation progressed, where it remained stable, and where it declined over time.

Apart from the student and school characteristics, also teacher characteristics were included as covariates in the analysis of student motivation. At the student level, the female students were more strongly motivated than the male students. Furthermore, at the teacher level, the students taught by teachers who had bachelor or master degrees were also found to be more highly motivated. And at the school level, students attending schools with high school prior achievement rates (8,1 – 9) also showed higher levels of motivation.

When controlling for these covariates and time linear, the students in experimental group 2 were on average found to be significantly more motivated. However, the interaction effect between time and the two interventions suggested a negative impact, which means that both experimental groups experienced a significant decrease in motivation over time. In other words, the student motivation in experimental group 2 was significantly higher than in the other groups, but the pattern declined over time. Consequently, the hypothesis on the effect of the interventions on student motivation, including the second one, was rejected.

6.4 Conclusion and Discussion

With respect to teaching quality, the findings in the multilevel analysis, especially with regards to the examination of the function of time (linear and quadratic), did not show significant effects. This implies that time could not be used to describe the changes in teaching quality over the three measurements. Moreover, the model fits

in these two models were not better than the empty model. The interaction effects between time and the interventions neither yielded the expected results, leading to the conclusion that both interventions had no effect on teaching quality improvement. The only significant effect found in the multilevel analysis, which was supported by a good model fit, was the effect of the second intervention as shown in Model 4 (Table 5.1); the teachers in experimental group 2 were found to be better teachers. However, following the results of the interaction effects, they did not improve more than the teachers in the other groups over the three measurements.

In conclusion, in terms of teaching quality, none of the two intervention groups improved more than the control group over the three measurements, which is why we had to reject the hypothesis on the effects of the interventions on teaching quality. Although teachers in experimental group 2 were better teachers they only improved their teaching quality during the first half of the intervention period. As regards student achievement, only the second intervention proved to have a significant effect, which became evident in the covariance and learning gain analyses. However, this effect could not be explained by an improvement in teaching quality. With respect to student motivation, both interventions showed a negative effect over time. Thus, only the combination of the elaborated standards and the teacher improvement program had produced a significant effect, but only in terms of student achievement. This effect could be considered as strong because it was confirmed by both the covariance and the learning gain analyses. Nevertheless, no effects were found in connection with the other variables, and therefore, overall the interventions did not yield the expected results.

Considering the above findings and conclusions, at least four questions could be raised: 1) What might explain a positive change of teaching quality in the second intervention group in the first half and a stability in the second half of the intervention program?, 2) What might explain the higher student achievement rates in the second intervention group?, 3) How could the interventions be improved in order to enhance the results? and 4) What are the directions for future research? First, we will address questions one till three and then we will also go into the limitations and strengths of this study. The last question will be dealt with in the final section.

6.4.1 Possible explanations for the positive change of teaching quality of the second intervention group from measurement one to two and a stability from measurement two to three

Both the descriptive (Table 4.1, chapter 4) as well as the multilevel analyses (Table 5.1, Model 4, chapter 5) revealed better teaching quality of the second intervention group compared to both the first intervention and the control groups. These findings could imply that when teachers participate in an effective teacher development program, their teaching quality will improve. Avalos (2011) for instance reviews publications in *Teaching and Teacher Education* over ten years (2000 – 2011) on teacher development programs and indicates that most studies show some form of impact of professional development on teachers' knowledge and practice. Specifically related to the effects of teacher development programs on teaching skill, Antoniou & Kyriakides (2013) have also reviewed a number of studies which support the links between teacher professional development and improvement in teaching skill.

However, teaching quality in the second intervention group in this study improved only from measurement one to two and was stable from measurement two to three. The materials presented during the intervention program might explain this finding. The first half of the program (introductory workshops and the first three monthly meetings) conducted before measurement two focused on the eight factors of the dynamic model. The content during this first half of the program likely facilitated teaching improvement in experimental group 2 from measurement one to two. In the second half of the program, the other three monthly meetings attempted to provide further chance for teachers to deepen their understanding on the elaborated standards document and the eight classroom factors of the dynamic model. This idea was accommodated through the development of lesson plans (2 meetings) and peer teaching (1 meeting), which were developed together with teachers in experimental group two due to their request. It has to be admitted that the last three meetings did not offer something new with regards to the classroom factors of the dynamic model, which could be the reason why no further improvement was found from measurement two to three.

6.4.2 Possible explanations for the higher student achievement rates in the second intervention group

We saw that after controlling for several background characteristics, the achievement rates of the students in the second intervention group were significantly higher than those of the students in the other groups, as demonstrated by their posttest scores and learning gain. Yet teaching quality improvement did not contribute to this result. So, how could this finding be explained?

It is possible that the achievement levels were stimulated by other changes took place in teaching quality, which were not measured in this study. Another possibility is that the initial higher motivation of the students in experimental group 2 enabled the teachers to work better, which in turn resulted in better student achievement outcomes.

More important is perhaps the design of the interventions. The combination of elaborated standards and a teacher improvement program, which involved a larger number of meetings, might have facilitated the teachers in having more control over their students' performance. It might explain why the first intervention, which only included the elaborated standards document and one day meeting, did not have a significant effect. This finding indicates that time and opportunity to learn do influence the results. As mentioned earlier, the review of Lauer, Snow, Martin-Glenn, Van Buhler, Soutemyer, and Snow-Renner (2005) addressed the importance of time in the professional development of teachers, arguing that in order to realize change, teachers need time and continuous support. This also applies to student achievement, which has been found to be closely related to the opportunity to learn (e.g. Scheerens, 2004). Therefore, it is possible that if more teacher sessions had been included in the first intervention, the outcome would have been better.

6.4.3 How could the interventions be improved?

Considering our results, which only showed an effect on student achievement, a critical evaluation of the design and the implementation of the interventions is in place. The quality of the results depends on the quality of the intervention, which includes content, methods, and intensity (Van der Werf, Creemers, De Jong, & Klaver, 2000). With respect to the design, a period of one school year might have

been too short for the teachers to significantly improve their methods of teaching. However, studies investigating developmental changes over time are quite scarce (Benner, 1984). Furthermore, as indicated earlier, teachers are generally reluctant to change and have a tendency to cling to their old ways (e.g. Thair & Treagust, 2003), which suggests that they need more time before they really gain an understanding the contents of interventions and are fully able to implement them. Moreover, several studies (e.g. Djalil, Jiyono, as cited in Jazadi, 2003; Kaluge, Setiasih, & Tjahjono, 2004; Utomo, 2005) have stressed the low teaching quality of Indonesian teachers. This group may therefore be given more time to learn to grasp the materials and to practice their implementation in the classes. Djalil (ibid) and Jiyono (ibid), for example, reported that teachers were simply not capable of using the relevant technology or adequately employing basic pedagogical tools, such as clarifying learning objectives, clearly explaining new concepts, giving examples, stimulating thinking through appropriate questioning and providing feedback on test results. Given this situation, it is possible that some of the materials introduced during the interventions, especially the second one, were too difficult for some teachers, who might have needed more time to learn to use them properly.

Next, as already mentioned, in the second intervention group both the student achievement rates and the teaching quality was better than in the other two groups, although there was no improvement effect in terms of teaching quality in intervention group 2. However, we are not sure whether the better teaching quality and student achievement rates in group 2 were the result of the combination of the elaborated standards and the teacher improvement program or of only the teacher improvement program. If an additional intervention condition had been added, for example, solely based on a teacher improvement program, we might have been able to determine more precisely which approach (the combination of the standards-based education and the teacher improvement program or only the teacher improvement program) was the most effective.

Concerning student motivation, our result was - although not expected - not very surprising. A number of previous studies as reviewed by Van Damme et al. (2006) have also indicated the minimum role of schools and teachers in influencing student motivation. 'Peer' has been found to be the most influential factor in increasing students' motivation to learn (Wong, 2007).

A more important issue in connection with the design, especially that of the second intervention, could be the staging of teaching quality as demonstrated by the dynamic model. In this respect, we refer to a study carried out in Cyprus to investigate whether teaching skills as represented by classroom factors and their dimensions can be classified into different stages. Using the Rasch model, this study identified five developmental stages of teaching skills, ranging from easy to more difficult (Kyriakides, Creemers, & Antoniou, 2009). The first three stages refer to direct and active teaching. During this trajectory the quantity (of factors) evolves into quality (of factors), which indicates that the quantity of teaching is a prerequisite for instruction. In addition, the skills are embedded in a gradual development pattern from more teacher-centered approaches to methods focused on an active involvement of the students. The last two stages are related to the differentiation dimension, which is a more demanding concept, as here teachers are expected not only to be aware of the different needs of their students but also to be capable of performing skills such as the appropriate application of tools, structuring, formulating questions, designing assessments for different groups of students (level 4) and using both direct/active teaching and the new teaching approach (level 5).

The staging method was not applied in this study because of the similarly low teaching quality of the teachers observed in the pilot studies. In order to make the training simpler, all teachers were provided with the same materials, although after each observation individual feedback was provided. It is possible that the teachers' perceptions of the difficulty of the materials differed; for some they may have been too difficult. It is also possible that had we used the staging method, improvement would have indeed been observed.

As regards the above factors, however, the teachers showed both progress and decline. In experimental group 2, for instance, the teachers succeeded in improving on all factors except questioning, whereas the teachers' performance in the other groups declined as regards three subscales, including questioning. In addition, the descriptive analysis in this study also indicated that questioning, assessment, and the creation of the classroom as a learning environment appeared to be easier activities for the teachers, because they had been practicing these skills from the beginning of the interventions. However, structuring, modeling, and

application might have been more difficult, given that on these factors the teachers scored low from the beginning of the intervention.

As previously indicated, the finding that the teachers participating in intervention 2 did not improve further during the last half of the intervention period might be influenced by the materials presented during this period. Despite the importance of the development of lesson plan and peer teaching as presented during the second half of the program, had it been known that structuring, modeling, and application were likely to be more difficult factors, they should be deepened in the last three meetings. In other words, it is important to identify the materials or the content of the program which are possibly difficult for teachers and then provide more sessions for those materials.

Furthermore, the implementation trajectory, which in this study was predominantly about the degree of transfer of what the teachers learned in their classrooms during the intervention period, is another issue that should also be taken into account. Closely related to this subject is fidelity, the assessment of which could help in gaining more insight into the strength of the implementation and the effect size of the interventions (Nelson, Cordray, Hulleman, Darrow, & Sommer, 2010). Whenever participants are offered to pick the intervention group of their own choice, higher levels of implementation and sustained practices are observed (Wehby, Maggin, Johnson, & Symons, 2010). Similarly, the better the implementation, the better the results are (Van der Werf et al. 2000). It has to be admitted that the degree of transfer of the content materials provided in the classroom interventions was not specifically measured in this study. The only measure used which could give some information on this matter was the list of attendees for each meeting, based on which we knew that only around half of the participants in experimental group 2 joined the meetings more or less consistently. It is therefore logical to assume that the low degree of transfer of the content of the interventions played a role in the unsatisfying results of this study.

Besides the design and the degree of transfer, another issue was that teachers have a tendency to rely (too) heavily on their materials and/or books, instead of focusing on what to do to improve both their teaching quality and the student outcome. This situation might have also influenced our findings. Unfortunately, not all textbooks

provide proper guidelines and relevant materials. In his critical review of two English course books, Priyanto (2009), for instance, concludes that generally no clear explanations are given of the use of language as a social process with a social purpose. Instead, textbooks tend to focus particularly on form, without explicating why and for which purposes learners should study this particular element. Another point Priyanto makes is that the books present the grammar separately; it is not integrated in the social context. In this way, no support is given to the learners in using the language as an instrument to achieve social goals. What Priyanto (ibid) does acknowledge, however, is that these particular books do provide the learners with a set of well-sequenced activities.

Furthermore, effective teacher improvement is known to require teacher guidance in the classroom. In this study, this was not possible due to practical reasons. Moreover, schools or other higher educational institutions, which have the authority to integrate these kinds of measures into their teacher improvement programs, were not involved in this project.

In sum, this subsection has identified several problems related to the design and the implementation of the interventions. These problems have indicated the limitations of the study at the theoretical level. Solving these issues is likely to lead to better intervention results. They include: the short time span of the interventions, the limited understanding of which approach might have yielded better student achievement rates and teaching quality in the second intervention group, the neglect of the developmental staging of teaching quality as introduced by the dynamic model, an insufficient focus on the degree of transfer in the implementation, the tendency of teachers to be text-book oriented, the absence of teacher guidance in the classroom, and the lack of school involvement. This study also has other limitations, which could be considered as more practical in nature. They will be presented in the following subsection.

One of the main limitations of this study was the modest reliability of the achievement test, especially in the pretest, despite its good reliability in the pilot study. One possible explanation is the different situations in the pilot and in the real study. In the pilot the students only completed the pretest whereas in the real study they did the pretest and filled in the questionnaire. Future research should therefore

make sure to conduct the pilot and the real study under the same conditions. In addition, if possible, it could be worth to test the instruments more than two times to make sure that all of their properties are satisfactory, although there is of course never a guarantee for this. Furthermore, 20 items could be considered as a small amount for an achievement test, which may have been another factor which influenced the low reliability. Had the items been doubled, then the reliability of the pretest and the posttest would have amounted to .68 and. 77, respectively, as indicated by the Spearman-Brown test.

However, considering the good reliability of the pilot study and the above arguments, the results of both the pretest and the posttest were retained in the analysis. Moreover, this decision was also supported by the literature. Ary, Jacobs, Sorensen, and Razavieh (2009) for instance, explain that the degree of reliability of a measure depends to a large extent on the use of the results: if the results are used to make a statement about a group or for research purposes, scores with only a modest reliability (coefficient between.50 to .60) are regarded as acceptable. In addition, although the test originated from a standardized test, it was modified and shortened for the specific purposes of this study, which is why it could be considered as either a researcher-made test or as a teacher-made test, the latter of which generally yields reliability scores of about .50 on average (Frisbie, 1998). According to Frisbie (ibid), this amount is tolerable for these types of tests. Thus, although it was a difficult decision, the use of the pretest results in this study had a sufficiently solid basis.

Secondly, although the total number of students who participated in the measurements was above 1,000 students, the sample loss was considerable, which was partly due to the fact that we could not visit the same class during the second measurement. Furthermore, in terms of schools and teachers, the number was rather limited, which may have had consequences for estimating the variability at the school and the classroom levels. Moreover, this study was limited to one type of school, namely madrasah, as well as to one subject and one grade. Future research should develop a way to maintain the number of participants throughout the measurements, and make sure that the observations are carried out in the same classes throughout all sessions. It is also considered desirable to study several subjects instead of just one, and organize a larger sample of participants. It would

of course also be preferable to examine a larger number of schools, although this would require more effort in terms of scheduling the visits for the different measurements.

Another limitation concerned the participants' assignment to the three groups. Although all participants voluntarily participated, the randomization in this study was considered successful because all of them had more or less similar background characteristics. However, eight schools in Serang (Banten) were all positioned in the control group because of the geographical location of this area. Thus the random assignment was not 100%. If possible, future research should avoid geographical barriers as a reason to exclude participants from joining the experimental groups. Instead, they should also be given access to the interventions.

In addition, although this research has been considered as a longitudinal study, three measurement points may not provide all the information necessary to make solid statements about whether and how the teachers learned from the interventions, and whether the materials introduced during the intervention period were applied adequately enough. Moreover, the three observations were conducted during the interventions, at which point the teachers might not yet have mastered the materials sufficiently enough to successfully use them in their own teaching and learning activities. It is possible that the students in the following years will benefit more from the interventions.

6.4.5 The strength of the study

In spite of the problems and limitations outlined above, one of the strengths of this study has been its experimental design, including the elements of EER. This approach has yielded theoretical-driven and evidence-based information based on which the educational improvement strategies might have led to better results. For instance, the results showed that the elaborated standards in combination with a teacher development program is a more effective intervention method, given its impact on the students' achievement. In addition, the longitudinal design has provided knowledge about how teaching quality and student outcomes can change during one school year.

With respect to the improvement strategies, the educational standards on their own, although clearly and specifically formulated, did not necessarily help the teachers improve their teaching quality, and neither did it prove to increase student achievement and student motivation. Although the standard-based education is aimed at the improvement of different factors on different levels, including teaching quality, significant effects on teaching quality and student outcome were not found in this study. On the other hand, the combination of the elaborated standards and a teacher improvement program did result in a better student achievement. In addition, the teaching quality of the teachers who participated in the second intervention was on average found to be higher. Moreover, teachers in this intervention group had a significant improvement in the first half of the intervention program. However, they did not improve further in the second half of the program, which made them did not improve more than the other groups over time and which implies several aspects including the fact that a period of one year including three points of measurement might not be enough.

Secondly, the study has shown the possibility of integrating EER into educational improvement initiatives, which has been advocated by a number of researchers (e.g. Creemers & Reezigt, 1996; Reynolds, Hopkins, and Stoll, 1993; Sammons, 1999). These studies have also demonstrated the complexity of such a combination. In our study, the classroom factors of the dynamic model, based on the results of the teacher effectiveness research, were used to promote the teachers' improvement in the second intervention program. Here the aim was to improve the student outcome through the enhancement of teaching quality. However, we saw an improvement of teaching quality only of the second intervention group in the first half of the program and no further improvement in the second half of the program. We also know that this group did not improve more in terms of teaching quality than the other groups over time, which could be related to the fact that the improvement only occurred in the first half of the program. Furthermore, we neither know whether the student achievement in experimental group 2 was the result of the combination of the elaborated standards and the teacher improvement program or only of the teacher improvement program. Thus, developing improvement measures is not as easy as finding factors that work for education and applying them. However, it is important

to recognize that the findings of EER are helpful in determining the direction of improvement.

Thirdly, different types of outcomes were measured at both the teacher and the student levels while controlling for different characteristics at the school, the teacher and the student levels. In addition, different sources of measurement were used, especially in measuring teaching quality: classroom observation by trained observers and a questionnaire on the students' perceptions of the teaching quality. At the student level, student achievement was tested via a pretest-posttest design while student motivation was measured during three points of measurement. In this way, not only cognitive but also non-cognitive outcomes were obtained.

6.5 Directions for future research

First of all, it is important to highlight that the elaborated standards on their own did not satisfy our expectations. The combination of the elaborated standards and a teacher improvement program appeared to be a stronger intervention, although only in terms of student achievement. However, as regards the study design attention should be paid to the aforementioned problems.

In view of future research, it is worth emphasizing the relevance of including the developmental stages of teaching quality. The study conducted in Cyprus by Antoniou (2009), which applied these stages as suggested by the dynamic model, resulted in improvements in both teaching quality and student achievement. The inclusion of this developmental stage may also address the problem of stability of the second intervention group after measurement two, which was likely to be related with the materials presented during the second half of the intervention program. Specific teaching skills in the next level can help focusing the content or the materials of the intervention. Especially in the context of Indonesia, it would be advisable to identify five developmental stages of teaching quality.

Secondly, as regards the question whether the better student achievement in experimental group 2 was the result of the combination of the elaborated standards and the teacher improvement program or only of the teacher improvement program, future research may add another group, which is only focused on a teacher improvement program. In addition, also more teachers and schools should be

included, although this approach would require more effort, while there is no guarantee that it would make the results easier to interpret.

Next, future research may opt for a longer period of intervention, containing more than three points of measurement. A longer intervention period is likely to better facilitate the teachers in taking in the information and use this knowledge in their classroom practices. In turn, more measurement points may enable researchers to map out the changes in teaching quality more adequately. Finally, as regards the low reliability of the pretest in this study, this issue should be tackled, because all instruments used should be sufficiently reliable.

With respect to the absence of teacher guidance in the classroom, follow-up research should include teacher supervision in implementing the content of the interventions. Furthermore, concerning the degree to which the content is implemented, instruments should be developed to enable the researchers to measure this extent. Finally, to further improve the overall results of future studies, the participation should be considered of both schools and other educational institutions in the interventions of the research.

Samenvatting

(The Dutch Summary)

Inleiding

Als reactie op de problemen in het onderwijs, met name de lage prestaties van leerlingen, heeft de Indonesische overheid in 2005 formeel onderwijsstandaarden ingevoerd als landelijke strategie om de kwaliteit van het onderwijs te verbeteren. Deze strategie komt overeen met de internationaal erkende opvatting dat het werken met standaarden, met name prestatiestandaarden, leidt tot verbetering van de leerprestaties van alle leerlingen ongeacht hun achtergrond, omdat ze voor leerkrachten duidelijk maken welke minimumdoelen ze bij leerlingen dienen na te streven (zie bijvoorbeeld Marzano, 1998; Marzano & Kendall, 1996a; Ravitch, 1995). Echter, tot op heden is er nog maar weinig bekend over de effecten van het werken met prestatiestandaarden, niet alleen in Indonesië maar ook in andere landen.

Op basis van studies in andere landen, met name in de Verenigde Staten, zijn twee zaken duidelijk geworden. In de eerste plaats is gebleken dat de effecten van standaarden zich vooral voordoen bij de betere leerlingen en in veel mindere mate bij leerlingen uit achterstandssituaties (zie bijvoorbeeld Lauer et al. 2005; Neumann et al. 2010, NRC, 2001; Thompson, 2009). In de tweede plaats doen zich tijdens de voorbereidings- en implementatiefase verschillende problemen voor die in zekere mate het achterwege blijven van de effecten van standaarden kunnen verklaren. Het eerste probleem is dat de standaarddocumenten in het algemeen erg algemeen, onduidelijk en vaag geformuleerd zijn (Choi et al. 2009; Gandall, 1996; Finn, Petrili, & Vanourek, 1998). In de tweede plaats zijn leerkrachten onvoldoende geschoold in het vertalen van de standaarden naar hun eigen lespraktijk. Deze situatie geldt ook in Indonesië. Ook daar zijn de standaarddocumenten geformuleerd in algemene termen waardoor ze voor leerkrachten weinig houvast bieden, en daarnaast krijgen leerkrachten maar weinig ondersteuning om de standaarden in praktijk te brengen (zie bijvoorbeeldHanafie, 2007; Lampung Post 2006).

De conclusie is dat er nog maar weinig bekend is over of en hoe onderwijsstandaarden de prestaties van leerlingen en dus de onderwijskwaliteit kunnen verhogen. Daarom is het voor beleidsmakers van belang om scholen te helpen bij het vinden van andere mogelijkheden om zichzelf te verbeteren. Het is relevant om hiervoor lering te trekken uit de resultaten van onderzoek naar onderwijseffectiviteit (Educational Effectiveness Research; EER), die theorie-

gestuurde en evidence-based informatie verschaffen over factoren op verschillende niveaus van het onderwijs die van invloed zijn op leerprestaties. Een groot aantal studies heeft aangetoond dat na controle voor achtergrondkenmerken van leerlingen, de leerkracht de grootste invloed heeft en dan met name diens vaardigheden in het geven van instructie, de leerstofkeuze en de nadruk op prestaties van leerlingen (zie bijvoorbeeld Creemers, 1994; Darling-Hammond, 1997; Doolaard, 1999; Fullan, 2001; Harris, 2002; Harris & Muijs, 2005; Marzano, 2007; OECD, 1994; Pilot, 2007; Van Der Werf, Creemers, De Jong, & Klaver, 2000).

Bovenstaande heeft geleid tot twee belangrijke conclusies. In de eerste plaats zijn prestatiestandaarden van belang die beschrijven wat alle leerlingen dienen te weten en te kunnen.Daartoe dienen de standaarddocumenten specifieke, duidelijke, concrete en operationele aanwijzingen te bevatten. In de tweede plaats is van belang dat leerkrachten ondersteuning krijgen bij het verbeteren van hun wijze van lesgeven, met name ten aanzien van de factoren waarvan is aangetoond dat ze effectief zijn voor de prestaties van leerlingen. In deze studie wordt verondersteld dat de combinatie van beide benaderingen – het werken met prestatiestandaarden en een leerkrachtverbeteringsprogramma – effectiever zal zijn voor verbetering van de kwaliteit van het onderwijs dan enkel de eerstgenoemde benadering.

Ten behoeve van het onderzoek zijn twee interventiesontwikkeld en vergeleken op hun effectiviteit in termen van verbetering van leerkrachtkwaliteit en leerprestaties, ten opzichte van een controlegroep waarbij geen sprake was van een interventie. Voor de interventies zijn twee ondersteunende documenten ontwikkeld. Het eerste bestond uit een duidelijke, specifieke en operationele uitwerking voor leerkrachten van de bestaande, nationale prestatiestandaarden. In het tweede document waren de kenmerken van effectief onderwijzen beschreven zoals naar voren komend uit EER, en in het bijzonder de acht factoren op klasniveau die deel uitmaken van het dynamisch model van onderwijseffectiviteit van Creemers en Kyriakides (2008). Het eerst genoemde document is gebruikt in beide interventies, en het tweede document – als aanvulling op het eerstgenoemde document - alleen in de leerkrachtverbeteringsinterventie.

Voor het toetsen van de effectiviteit van de twee interventies is gedurende een schooljaar een experimenteel onderzoek uitgevoerd, waarin drie metingen hebben

plaatsgevonden in zowel de twee experimentele groepen als in de controlegroep. De drie hoofdvragen van het onderzoek waren: 1) wat zijn de effecten van de interventie met de uitgewerkte standaarden en van de interventie met de standaarden in combinatie met het leerkrachtverbeteringsprogramma op de leerkrachtkwaliteit?; 2) wat zijn de effecten van de twee genoemde interventies op de leerprestaties van de leerlingen?; 3) in welke mate worden de effecten van de interventies op de leerlingprestaties verklaard door de verbetering van de leerkrachtkwaliteit?

In lijn met deze onderzoeksvragen zijn drie hypothesen geformuleerd. De eerste hypothese was dat leerkrachten die door middel van het uitgewerkte standaarddocument duidelijk inzicht krijgen in de doelen die leerlingen dienen te bereiken (interventie 1), hun kwaliteit van lesgeven meer verbeteren en hogere prestaties bij leerlingen bereiken dan leerkrachten die enkel de originele nationale standaarddocumenten ter beschikking hebben (controlegroep). De tweede hypothese was dat leerkrachten die, naast het hebben van een duidelijk en specifiek standaarddocument, tevens ondersteund worden met een leerkrachtverbeteringsprogramma (interventie 2), meer hun kwaliteit van lesgeven verbeteren en hogere prestaties bij leerlingen bereiken, in vergelijking tot de leerkrachten die enkel de nationale documenten hebben (controlegroep), als ook in vergelijking tot de leerkrachten die het uitgewerkte standaarddocument hebben gekregen (interventie 1). De derde hypothese was dat de eventueel gevonden effecten van beide interventies op de prestaties van de leerlingen verklaard kunnen worden door de verbetering van de kwaliteit van het lesgeven van de leerkrachten.

Opzet van het onderzoek

Het onderzoek is uitgevoerd in Indonesië en was gericht op het vak Engels (begrijpend lezen) op madrasahTsanawiyah scholen voor Junior Voortgezet Onderwijs in twee provincies: DKI Jakarta en Banten. Dit zijn alle Moslimscholen die vallen onder het Indonesische Ministerie van Religieuze Zaken. De deelnemers (59 leraren, 57 scholen, 2431 leerlingen in het tweede leerjaar)namen vrijwillig deel aan het onderzoek en zijn random toegewezen aan een van de twee experimentele condities dan wel aan de controlegroep. De leerkrachten in de eerste experimentele

groep namen deel aan interventie 1, waarin het standaarddocument “the elaboratedstandards of English” ter beschikking werd gesteld en toegelicht in een workshop gedurende een dag.De leerkrachten in de tweede experimentele groep namen deel aan interventie 2. In deze interventie werd tevens het uitgewerkte standaarddocument beschikbaar gesteld en toegelicht in een workshop van een dag (gezamenlijk met de leerkrachten die deelnamen aan interventie 1). Daarnaast kregen de leerkrachten in deze groep het document “becominganeffective teacher of English”, waarin strategieën waren beschreven voor het verbeteren van de kwaliteit van lesgeven op basis van de klasfactoren uit het dynamisch model van onderwijseffectiviteit. Deze groep ontving tevens een workshop van een dag over effectief onderwijzen. Daarnaast werdenvoor deze groep zes bijeenkomsten van een halve dag georganiseerd,waarin werd besproken hoe de standaarden en de klasfactoren in de dagelijkse praktijk konden worden toegepast.

De effecten van de interventies zijn vastgesteld door middel van het meten van de kwaliteit van lesgeven door de leerkrachten –tevens een intermediërende variabele (zie hypothese 3) en de leerprestaties en prestatiemotivatie van de leerlingen. De kwaliteit van lesgeven is drie maal gedurende het schooljaar gemeten – begin, midden en eind – met behulp van een observatie-instrument, toegepast door onafhankelijke observatoren, evenals door middel van een leerlingvragenlijst die twee maal is afgenomen (midden en eind schooljaar).De leerprestaties van de leerlingen zijn twee maal gemeten (begin en eind schooljaar) met een toets voor begrijpend lezen en de prestatiemotivatie is drie maal (begin, midden en eind schooljaar) gemeten met een vragenlijst.

Samenvatting van de resultaten

De onderzoeksgegevens zijn allereerst beschrijvend geanalyseerd, waarin gekeken is naar verschillen tussen de drie groepen leerkrachten in de kwaliteit van lesgeven en de leerresultaten en prestatiemotivatie van de leerlingen. Tevens zijn correlaties berekend met de achtergrondkenmerken van leerlingen, leerkrachten en scholen. Vervolgens zijnmultilevel analyses uitgevoerd ten behoeven van het beantwoorden van de onderzoeksvragen en het toetsen van de hypothesen. De resultaten worden hieronder samengevat:

Vraagstelling 1: Omdat de kwaliteit van lesgeven drie keer is gemeten zijn de gegevens geanalyseerd met een multilevelgroeicurve model (meetmomenten, genest binnen leerkrachten). De resultaten laten zien dat de variantie in de kwaliteit van lesgeven tussen de verschillende meetmomenten ongeveer even groot is als de variantie tussen leerkrachten. Er is geen significant effect van tijd en tevens zijn er geen verschillen tussen leerkrachten in toename van de kwaliteit van lesgeven over de tijd. Tenslotte, en het meest belangrijk voor de hypothese, blijkt dat de leerkrachten in beide experimentele condities niet meer vooruitgaan over de drie meetmomenten dan de leerkrachten in de controleconditie, en tevens dat de leerkrachten in de tweede experimentele conditie niet meer vooruitgaan dan de leerkrachten in de eerste experimentele conditie. Hypothese 1 wordt dus niet bevestigd door de resultaten. Wel laten de beschrijvende resultaten zien dat de leerkrachten in de tweede experimentele conditie hun kwaliteit van lesgeven hebben verbeterd tussen de eerste en de tweede meting. Daarna zijn ze stabiel gebleven.

Vraagstelling 2: Voor het toetsen van de tweede hypothese betreffende de leerprestaties is zowel een co-variantieanalyses als een leerwinstanalyse uitgevoerd. Er is sprake van een tweeniveaumodel waarin leerlingen zijn genest binnen leerkrachten. In beide analyses zijn achtergrondkenmerken van leerlingen (sekse en SES) en scholen (provincie en de eerdere schoolscore op het nationaal examen Engels) die uit de beschrijvende analyses als relevant naar voren kwamen, opgenomen als controlevariabelen, en daarnaast zijn in de co-variantieanalyse tevens de scores op de pre-test als co-variaat opgenomen. Uit de resultaten blijkt dat de leerlingen in de tweede experimentele conditie, zowel in het co-variantie als in het leerwinstmodel hogere resultaten behalen dan de leerlingen in de controlegroep. Echter er is geen significant effect van de eerste experimentele conditie. Hypothese 2 betreffende de leerprestaties is hiermee slechts bevestigd voor zover deze betrekking heeft op de experimentele conditie 2 (de standaarden en het leerkrachtverbeteringsprogramma).

Voor het toetsen van hypothese 2 betreffende de prestatiemotivatie van de leerlingen is een multilevel groeicurve model geschat, waarin de drie meetmomenten zijn genest binnen leerlingen. De resultaten laten zien dat in beide experimentele groepen sprake is van een hogere mate van afnemende motivatie dan in de

controlegroep, en dus wordt hypothese 2 met betrekking tot de motivatie van leerlingen niet ondersteund.

Vraagstelling 3: Met het oog op het toetsen van de derde hypothese is de variabele 'verbetering van de kwaliteit van lesgeven' toegevoegd aan het co-variantie- en het leerwinstmodel (zie hierboven). Deze variabele bleek niet significant samen te hangen met de leerresultaten van de leerlingen op de posttest noch met de leerwinst tussen pre- en posttest, zodat hypothese 3 niet werd ondersteund.Met andere woorden, het gevonden effect van de experimentele conditie 2 op de verbetering van de leerprestaties van de leerlingen wordt niet verklaard door verbetering van de kwaliteit van lesgeven in deze experimentele conditie.

Conclusies en discussie

De conclusies van het onderzoek zijn 1) dat er geen effect is gevonden van de twee interventies op de verbetering van de kwaliteit van lesgeven door leerkrachten, maar dat er in de tweede interventiegroep wel een verbetering is opgetreden tussen het eerste en tweede meetmoment; 2) dat alleen de tweede interventie een positief effect heeft op de verbetering van de leerprestaties; 3) dat dit effect niet verklaard kan worden door de verbetering van de kwaliteit van lesgeven; en 4) dat beide interventies hebben geleid tot een grotere afname van de motivatie van leerlingen. Kortom, alleen de combinatie van uitgewerkte prestatiestandaarden en het programma voor verbetering van leerkrachtgedrag heeft geleid tot de gewenste resultaten in termen van verbetering van de leerprestaties. Dit effect was zichtbaar in zowel de co-variantie als in de leerwinstbenadering. Echter, het effect van deze interventie op de verbetering van de leerprestaties is niet toe te schrijven aan een verbetering van de kwaliteit van lesgeven door de leerkrachten. Deze conclusie leidt tot vier belangrijke vragen. De eerste vraag betreft een mogelijke verklaring voor de bevinding dat er in de tweede interventiegroep wel een verbetering is opgetreden tussen het eerste en het tweede meetmoment, maar daarna niet meer. Deze verklaring kan gezocht worden in de inhoud van het trainingsprogramma. De eerste helft van de training (introductie workshop en de eerste drie maandelijkse bijeenkomsten) vond plaats voor het tweede meetmoment, en was vooral gericht op de introductie van het uitgewerkte standaardendocument en de acht factoren van het dynamisch model. In de tweede helft, de drie andere maandelijkse

bijeenkomsten, werd vooral gewerkt aan verdere verdieping van het begrip van leerkrachten van deze aangeboden informatie. Wellicht verklaart het feit dat in deze tweede helft van de interventie geen nieuwe informatie meer werd aangeboden waarom zich na het tweede meetmoment geen verdere verbetering meer heeft voorgedaan in de kwaliteit van lesgeven.

De tweede vraag betreft de mogelijke verklaringen voor de hogere prestaties in de tweede interventiegroep. Tenminste drie verklaringen zijn aannemelijk: 1) er was sprake van andere dan de gemeten verbeteringen in de kwaliteit van lesgeven die tot betere leerprestaties hebben geleid; 2) de leerlingen in de tweede interventie waren, ondanks hun afnemende patroon, meer gemotiveerd, zodat leerkrachten gemakkelijker hun werk konden doen en hogere prestaties konden realiseren; en 3) het feit dat er in de tweede interventie meer bijeenkomsten waren hebben op zichzelf geleid tot hogere prestaties van leerlingen.

De derde vraag heeft betrekking op de aspecten in de interventiesdie verbetering behoeven zodat ze leiden tot betere resultaten. Gedurende de studie zijn diverse problemen naar voren gekomen die geïdentificeerd kunnen worden als beperkingen van het onderzoek. Deze problemen betreffen de korte looptijd van de interventies en de korte periode tussen de meetmomenten. Een ander probleem is het feit dat het onzeker is of het effect van alleen de tweede interventie is toe te schrijven aan de combinatie van de uitgewerkte standaarden en het trainingsprogramma voor leerkrachten dan wel aan het feit dat deze interventie meer bijeenkomsten omvatte dan de eerste interventie en dus leerkrachten meer bewust bleef houden van het monitorenvan de prestaties van hun leerlingen.

De vierde vraag betreft de implicaties voor toekomstig onderzoek. In het huidige onderzoek is geen aandacht besteed aan de ontwikkelingsstadia in de kwaliteit van lesgeven, zoals onderscheiden in het dynamisch model. Alle leerkrachten in de tweede interventie kregen dezelfde training en materialen. Echter, het is mogelijk dat voor sommige leerkrachten het trainingsaanbod te moeilijk was en voor anderen juist te gemakkelijk. In toekomstig onderzoek zou de training meer moeten worden aangepast aan het niveau van individuele leerkrachten. Een ander punt is dat meer aandacht besteed zou moeten worden aan de transfer van het geleerde tijdens de training naar de dagelijkse klaspraktijk, mede ook omdat leerkrachten zeer geneigd

zijn te leunen op het materiaal en de tekstboeken die ze hebben en veel minder vertrouwen op wat ze geleerd hebben met betrekking tot de verbetering van hun lesgedrag. Tenslotte is ook gewenst dat alle leerkrachten daadwerkelijk alle trainingsbijeenkomsten bijwonen, en wellicht ook dat de trainingsperiode wordt verlengd. Met betrekking tot het design van de studie, is het wenselijk om een extra interventie op te nemen, waarin alleen sprake is van een trainingsprogramma voor leerkrachten, zonder het beschikbaar stellen van een uitgewerkt standaarddocument. Dit, om beter zicht te krijgen op de vraag of de effecten van de tweede interventie zijn toe te schrijven aal het trainingsprogramma voor leerkrachten, dan wel aan de combinatie daarvan met het standaardendocument. Ook verdient het aanbeveling om nauwkeuriger en vaker na te gaan of leerkrachten daadwerkelijk het geleerde tijdens de trainingsbijeenkomsten toepassen in hun lespraktijk.

Ondanks de hierboven genoemde beperkingen, heeft de studie bijgedragen aan de kennisbasis met betrekking tot het verbeteren van de kwaliteit van onderwijs door te laten zien dat de combinatie van uitgewerkte leerstandaarden en een trainingsprogramma voor leerkrachten gericht op de verbetering van de kwaliteit van lesgeven inderdaad leidt tot hogere prestaties van leerlingen, en dat alleen het aanbieden van uitgewerkte standaarden niet tot de gewenste effecten leidt. Deze bevinding bevestigt opnieuw dat de leerkracht de meest bepalende factor is voor het verbeteren van de leerprestaties van leerlingen. Verder onderzoek is nodig om te bepalen welke factoren betreffende de kwaliteit van lesgeven door leerkrachten verantwoordelijk zijn voor dit effect en op welke wijze deze factoren kunnen worden bevorderd in trainingsprogramma's voor leerkrachten.

REFERENCES

Althoff, S. E., Linde, K. J., Mason, J. D., Nagel, N. M., & O'Reilly, K. A. (2007). *Learning objectives: Posting & communicating daily learning objectives to increase student achievement and motivation.Saint Xavier University & Pearson Achievement Solutions*, 1-76.

American Federation of Teachers (2009). *The instructional demands of standards-based reform.* Washington DC: American Federation of Teachers.

Antoniou, P. (2009). *Using the dynamic model of educational effectiveness to improve teaching practice*: *building an evaluation model to test the impact of teacher professional development programs*. Unpublished PhD thesis. University of Cyprus: Department of Education.

Antoniou, P., Demetriou, D. & Kyriakides, L. (2006). *Towards a Dynamic Model of Educational Effectiveness. A Meta-analysis of Studies Investigating the Impact of School Factors on Student Achievement Gains. Proceedings of the IX Conference of the Cyprus Pedagogical Association. Nicosia*.

Antoniou, P. & Kyriakides, L. (2013). *A dynamic integrated approach to teacher professional development: Impact and sustainability of the effects on improving teacher behavior and student outcomes. Teaching and Teacher Education*, 29: 1 – 12.

Apthorp, H. S., Dean, C. B., Florian, J. E., Lauer, P. A., Reichardt, R.,& Snow-Renner, R. (2001). *Standards in classroom practice: Research synthesis*. Washington D.C.: Mid-Continent Research for Education and Learning.

Asian Development Bank. (2006). *Indonesia: madrasah education development project.* Project Number: 37475-01. Jakarta: SMEC International Pty Ltd.

Askew, M. & William, D. (1995). *Recent research in Mathematics education 5-16*. London: Office for Standards in Education, 53.

Ary, D., Jacobs, L. D., Razavieh, A., & Sorensen, C. (2006). *Introduction to research in education*. Belmont, CA: Thomas Higher Education.

Avalos, B. (2011). *Teacher professional development in Teaching and Teacher Education* over ten years. *Teaching and Teacher Education*, 27: 10 – 20.

Badan Pusat Statistik. (2011). *Statistik Indonesia: statistical yearbook of Indonesia 2011*. Jakarta: Badan Pusat Statistik.

Badan Standar Nasional Pendidikan. (2010). *Statistik*. Retrieved from http://bsnp-indonesia.org/id/?page_id=39/.

Baines, L. A., & Stanley, G. K. (2006). *The iatrogenic consequences of standards-based education. The Clearing House,* 79(3), 119-123.

Baker, L., & A. Wigfield. (1999). *Dimensions of children's motivation for reading and their relations to reading activity and reading achievement. Reading Research Quarterly 34*(4): 452–77.

Benner, P. (1984). *From novice to expert: Excellence and power in clinical nursing practice*. San Francisco: Addison-Wesley.

Bennett, N., Desforgess, C., Cockburn, A., & Winkelson, B. (1981). *The quality of pupil learning experiences* (interim report). Lancaster, UK: University of Lancaster, Centre for Educational Research and Development.

Block, A.W. (1985)(2nd printing). *Effective schools: A summary of research*. Arlington, VA: Educational Research Service.

Bloom, H.S. (2006). *The core analytics of randomized experiments for social research*. (MDRC working papers on research methodology. Available at http://www.mdrc.org/publications/437/full.pdf.

Bohn, M., Roehrig, A.D., & Pressley, M. (2004). *The first days of schools in the classrooms of two more effective and four less effective primary-grade teachers. The Elementary School Journal, 104* (4).

Borich, G.D. (1992). *Effective teaching methods,* 2nd ed. New York: Macmillan.

Bosker, R.J. & Scheerens, J. (1994). *Alternative models of school effectiveness put to the test. International Journal of Educational Research, 21* (2), 159–180.

Brophy, J.E. (1981). *Teacher praise: A functional analysis. Review of Educational Research 51*(1): 5-32.

Brophy, J.E. & Good, T.L. (1986). Teacher behavior and student achievement. In Wittrock , M.C. (ed.), *Handbook of research on teaching* (3rd edn, pp. 328 - 375). New York: Macmillan.

Burke, L. M. & Marshall, J. A. (2010). *Why national standards won't fix American education: Misalignment of power and incentives. backgrounder. no. 2413.* Heritage Foundation.

Buttram, J .L. & Waters, J. T. (1997, September). *Improving America's schools through standards-based education. NASSP Bulletin, 81*(590), 1–6.

Case, R. (1993). Theories of learning and theories of development. *Educational Psychologist, 28* (3): 46-50.

Cahyono, B.Y. (2008). *The continuous improvement learning programme for Englishfor English language teachers: An Indonesian experience. In*

M.E. Vethamani& M.K. Kabilan (Eds.), Practices and issues in English language teacherdevelopment (61-78). Petaling Jaya, Malaysia: Sasbadi Sdn. Bhd.

CEQDA (2007). *Laporan dampak kebijakan akreditasi, BOS, dan sertifikasi terhadap masa depan madrasah di DKI Jakarta.* Jakarta: Kerjasama LAPIS dan CEQDA UIN Jakarta.

Chalker, D. M., & Haynes, R. M. (1994). *World class schools: New standards for education*. Lancaster, PA etc.: Technomic.

Chambers, L. & Dean, C. (2000).*Noteworthy perspectives on implementing standards-based education.*Washington D.C.: Mid-Continent Research for Education,and Learning.

Chodijah, I. (2012). *Sekolah lembaga tempat mendidik atau mengeluarkan ijazah saja?* Retrieved online from http://edukasi.kompasiana.com/2012/04/18/sekolah-lembaga-tempat-mendidik-atau-mengeluarkan-ijazah-saja/.

Choi, D., de Vries, H. J., & Kim, D. (2009). *Standards education policy development: Observations based on APEC research*. Rotterdam; Rotterdam: Erasmus Research Institute of Management, Erasmus University; Erasmus University Rotterdam Host.

Cizek, G.J. (2001). *Setting performance standards: concepts, methods, and perspectives*. Mahwah, NJ: Lawrence Erlbaum Associates

Cohen, J. (1992). A Power primer. *Psychological Bulletin, 112*(1), 155-159.

Coleman, J.S., Campbell, E.Q., Hobson, C.F., McPartland, J., Mood, A.M., Weinfeld, F.D., & York, R.L. (1996). *Equality of educational opportunity*. Washington, DC: US Government Printing Office.

Corcoran, T. (1995). *Transforming professional development for teachers: A guide for state policymakers*. Washington, DC: National Governor's Association.

Cotton, K. (1995). *Effective student practices: A research synthesis*. Available at http://www.nwrel.org/scpd/esp/esp95.htm1#1.

Creemers, B.P.M. (1994). *The effective classroom.* London: Cassell.

Creemers, B. & Kyriakides, L. (2008). *The Dynamics of Educational Effectiveness.A contribution to Policy, Practice and Theory in Contemporary schools.* New York & London: Routledge Taylor & Francis Group.

Creemers B.P.M., & Reezigt G.J. (1996). *School level conditions affecting the effectiveness of instruction. School Effectiveness and School Improvement*, 7 (3), 197-228.

Creemers, B.P.M. & Scheerens, J. (n.d). *Developments in educational effectiveness research programme*. Available at http://doc.utwente.nl/26843/1/Creemers94developments.pdf.

Creemers, B.P.M., Stoll, L., Reezigt, G., and the ESI team. (n.d). *Effective school improvement – ingredients for success: The results of an international comparative study of best practice case*. Available at http://www.rug.nl/staff/b.p.m.creemers/Effective%20School%20Improvement%20Ingredients%20for%20Success.pdf.

Darling-Hammond, L. (1997). *Doing what matters most: investing in quality teaching*. New York: National Commission on Teaching & America's Future.

Darling-Hammond, L. (2000). Teacher quality and student achievement: A review of state policy evidence. *Educational Policy Archives, 8*(1). Retrieved from http://epaa.asu.edu/epaa/v8n1.

De Corte, E. (2000). *Marrying theory building and the improvement of school practice: a permanent challenge for instructional psychology*. *Learning and Instruction, 10* (3): 249-266.

De Jong, R., Westerhof, K.J., & Kruiter, J.H. (2004). *Empirical evidence of a comprehensive model of school effectiveness: A multilevel study in mathematics in the 1st year of junior general education in the Netherlands*. *School Effectiveness and School Improvement, 15*(1), 3-31.

Delandshere, G., & Petrosky, A. (2004). *Political rationales and ideological stances of the standards-based reform of teacher education in the US. Teaching & Teacher Education: An International Journal of Research and Studies, 20*(1), 1-15.

Departemen Pendidikan Nasional (2008). *Ikhtisar Data Pendidikan Nasional Tahun 2007/2008*. Jakarta: Badan Penelitian dan Pengembangan Depdiknas.

Doherty, K.M. (2001, January 11). *Poll: Teachers support standards with hesitation*. *Education Week / Quality Counts* 20(17):20.

Donnelly, K. (2007). *Australia's adoption of outcome based education: A critique*. *Issues in Educational Research, 17*(2): 183-205.

Doolaard, S. (1999). *Schools in change or schools in chain?: the development of educational effectiveness in a longitudinal perspective*. Enschede: Twente University Press.

Dowson, M., McInerney, D. M., & Van Etten. (2007). *The state of play in standards and standards reform.* In McInerney, D. M., Van Etten, S., & Dowson, M. *Standards in education*. Charlotte, NC: Information Age Publishing.

Doyle, W. (1986). *Classroom organization and management.* In Wittrock , M.C. (ed.), *Handbook of research on teaching* (3rd edn, pp. 392-431). New York: Macmillan.

Echevarria, J., Short, D., & Powers, K. (2006). *School reform and standards-based education: a model for English-language learners*. Heldref Publications.

Education Commission of the States (2002). *Standard-based education. No child left behind issue brief*. Washington D.C.: Department of Education.

Education, Audiovisual and Culture Executive Agency (2011). *Teaching Reading in Europe: Context, policies, and practices.* Brussel: Education, Audiovisual and Culture Executive Agency.

Elliott, J. L., Thurlow, M. L. (1997).*Opening the door to educational reform: Understanding standards.* Minneapolis, MN: National Center on Educational Outcomes.

Everston, C.M., Anderson C., Anderson, L., & Brophy, J. (1980). *Relationships between classroom behavior and student outcomes in junior math and English classes. American Educational Research Journal*, *17*, 46-60.

Faizi, W., Shakil, A. F., & Lodhi, F. A. (2011). *The main reasons of declining educational standards at secondary level in Karachi, Pakistan. Educational Research and Reviews, 6*(2), 152-160.

Finn, C.E., Jr., Petrilli, M.J., & Vanourek, G. (1998). *The State of State Standards.*Washington, D.C: Thomas B. Fordham Foundation.

Fleischman, H. L., Hopstock, P. J., Pelczar, M. P., Shelley, B. E. (2010). *Highlights from PISA 2009: Performance of U.S. 15-year-old students in reading, mathematics, and science literacy in an international context. NCES 2011-004.* US Department of Education.

Florian, J. (1999). *Teacher survey of standards-based instruction: Addressing time.*Aurora, CO: Mid-Continent Research for Education and Learning.

Fraser, K. (ed) (1996). *What will it take?*: *standards-based education reform for all students*. Alexandria, VA: Centre for Policy Research on the Impact of General and Special Education Reform.

Frisbie, D. A. (1998). *Reliability of scores from teacher-made tests*. *Educational Measurement: Issues and Practice, 10,* 55-65.

Fullan, M. (2001). *The new meaning of educational change*. London: Routledge Falmer.

Gandal, M. (1996). *Making standards matter, 1996: an annual fifty state report on efforts to raise academic standards*. Washington, DC: American Federation of Teachers.

Gibbs, T.J. & Howley, A. (2000). *"World-class standards" and local pedagogies: Can we do both?*. *Threshold in Education* (Aug/Nov 2001). Retrieved online from http://m.cedu.niu.edu/lepf/foundations/thresholds/journal/2001. Volume. XXVII/Issue.3_4/51.World-Class. Standards.and.Local.Pedagogies. Can. We. Do.Both.pdf.

Grabe, W. (2009). *Reading in a second language; moving from theory to practice*. CambridgeUniversity Press.

Goldstein, H. (2003). *Multilevel statistical models*. London: Hodder Arnold.

Goodwin, B. (2003). *Digging deeper: Where does the public stand on standards-based education? Issues brief*. Aurora, CO: Mid-Continent Regional Educational Lab.

Gustafsson, J.E. (2010). Longitudinal design. In Cremeers, B.P.M., Kyriakides, L., & Sammons, P. (eds). *Methodological advances in educational effectiveness research: quantitative methodology series*. Oxon: Routledge.

Hamilton, L. S., Stecher, B. M., Yuan, K. (2008). *Standards-based reform in the United States: History, research, and future directions.* Washington, DC: Center on Education Policy.

Hammer, D. (1998). *The standards teacher. Standards for excellence in education.*Washington D.C.: Council for Basic Education.

Hanafie, I. (2007). *Plus minus kurikulum tingkat satuan pendidikan (KTSP).* Retrieved online from http://meysiska.blogspot.com/2009/05/plus-minus-kurikulum-tingkat-satuan.html, on November 13, 2009.

Harris, A. (1998). Effective teaching: a review of the literature. *School Leadership & Management*, *18* (2), 169-183.

Harris, A. (2002). *School improvement: what's in it for school?*. London: Falmer Press.

Harris, A. &Muijs, D. (2005). *Improving schools through teacher leadership*. London: Open University Press.

Heck, R. H. (2007). *Examining the Relationship Between Teacher Quality as an Organizational Property of Schools and Students' Achievement and Growth Rates. Educational Administration Quarterly, 43* (4), 399-342.

Heck, R.H. & Marcoulides, G.A. (1996). *School culture and performance: Testing the invariance of an organizational model. School Effectiveness and School Improvement, 7* (1): 76 – 95.

Hendayana, S. (2007). *Development of INSET model for improvinf teacher professionalism in Indonesia. NUE Journal of International Educational Cooperation*, Volume 2, p. 97 – 106.

Hill, H.C., (2001). *Policy is not enough: Language and the interpretation of state standards. American Educational Research Journal* (38), 289-318.

Hill, P.W. & Rowe, K.J. (1996). *Multilevel modeling in school effectiveness research. School Effectiveness and School Improvement, 7* (1), 1-34.

Hoyle, C. D., O'Dwyer, L. M., &Chang, Q. (2011). *How student and school characteristics are associated with performance on the Maine high school assessment. Issues & Answers, REL 2011- No. 102.*

Hunt Jr., J.B, Rizzo, J.A, & White, A.D. (2009). *Realizing the promise of standards - based reform. The Hunt Institute's BLUEPRINT for Education Leadership* No. 3. Retrieved from http://www.hunt-institute.org/elements/media/files/Blueprint-Number-3-June-2009.pdf.

Huriyah, L. (2009). *Problematika Sertifikasi Guru. Lentera*¸No 13 Volume 8.

Hustinx, P. W. J., Kuyper, H., Van, der Werf, & Dijkstra, P. (2009). *Achievement motivation revisited: New longitudinal data to demonstrate its predictive power. Educational Psychology, 29*(5), 561-582.

Indonesia Education Act No 20 Year 2003.

Irons, E.J., Carlson, N.L., Lowery-Moore, H., Farrow, V.R. (2007). *Standards and Accountability Implementation, Why, How, Where: Teachers' Perceptions. Journal of Educational Research and Policy Studies*, volume 7 No 2.

Janson, C. G. (1990). Retrospective data, undesirable behaviour, and the longitudinal perspective., in D. Magnusson & L.R. Bergman (eds) *Data Quality in Longitudinal Research*. Cambridge: Cambridge University Press.

Jazadi, I. (2003). *An investigation of current constraints and potential resources for developing learner-centred curriculum frameworks for English language at high schools in Lombok, Indonesia. Unpublished PhD thesis*. Adelaide, Australia: Division of Education, Arts and Social Sciences, University of South Australia.

Jencks, C., Smith, M., Acland, H., Bane, M.J., Cohen, D., Gintis, H., Heyns, B., & Michelson, S. (1972). *Inequality: a reassessment of the effects of family and schooling in America.* New York: Basic Books.

Jo, B. & Muthen, B. (2003). Longitudinal studies with intervention and noncompliance: Estimation of causal effects in growth mixture modeling. In Reise, S.P. & Duan N. (eds). *Multilevel modeling: Methodological advances, issues, and applications*, pp. 112-139. Mahwah, NJ: Lawrance Erlbaum Associates.

Joyce, B., Weil, M. & Calhoun, E. (2000). *Models of Teaching*. Boston: Allyn and Bacon.

Kaluge, L., Setiasih, Tjahjono, H. (2004). *The Quality Improvement of Primary Children Learning through a School-Based Programme in Indonesia*. A Research Paper. East Java: Universitas Surabaya.

Kane, T.J., Taylor, E.S., Tyler, J.H. & Wooten, A.L. (2010). *Identifying effective classroom practices using student achievement data*. *The Journal of Human Resources 46*(3): 587 – 613.

Kompas (2008, May 17). *Kualitas SD masih diabaikan: Delapan standar belum dipenuhi*. Retrieved online from http://transparansipendidikan.blogspot.com/2008/05/kualitas-sd-masih-diabaikan-delapan.htm.

Kompas (2010, May). *Foke prihatin, 39.179 siswa SMP gagal UN*. Retrieved online from http://edukasi.kompas.com/read/2010/05/06/16552537/.

Kompasiana (2011, January 30). *Indonesia peringkat 10 besar terbawah dari 65 negara peserta PISA*. Retrieved online from www.edukasi.kompasiana.com/.../**indonesia**-**peringkat**-10-besar-terbawah-dari-65- negara-peserta-**pisa**.

Kulik, J. A., and C.-L. Kulik. (1989). "Effects of Ability Grouping on Student Achievement." *Equity and Excellence* 23, 1-2: 22-30.

Kyriakides, L. (2005). Extending the comprehensive model of educational effectiveness by an empirical investigation. *School Effectiveness and School Improvement,* 16(2), 103-152.

Kyriakides, L., Campbell, R.J., & Gagatsis, A. (2000). The significant of the classroom effect in primary school: An application of Creemers comprehensive model of educational effectiveness. *School Effectiveness and School Improvement, 11*(4)*,* 501-529.

Kyriakides, L. & Creemers, B.P.M. (2006). Using the dynamic model of educational effectiveness to introduce a policy promoting the provision of equal opportunities to students of different social groups. In D.M. McInerney, S. Van Etten, & M. Dowson (eds), *Research on sociocultural influences on motivation and learning*, vol. 6: *Effective schooling*. Greenwich CT: Information Age Publishing.

Kyriakides, L., Creemers, B.P.M. & Antoniou, P. (2009). Teacher behaviour and student outcomes: Suggestions for research on teacher training and professional development. *Teaching and Teacher Education, 25* (1), 12-23.

Lampung Post (6 Desember 2006). Kurikulum: pemahaman KTSP bervariasi. Retrieved online from http://www.lampungpost.com/cetak/berita.php?id=2006120601353359 on 2009, November 13.

Lauer, P.A., Snow, D., Martin-Glenn, M., Van Buhler, R.J., Soutemyer, K., & Snow-Renner R., (2005). *The influence of standards on K-12 teaching and student learning: A research synthesis*. Regional Educational Laboratory.

Lavy, V. (2011). *What makes an effective teacher? Quasi-experimental evidence*. Working paper 16885. Cambridge, MA: National Bureau of Economic Research.

Leinhardt, G., Weidman, C., & Hammond, K.M. (1987). Introduction and integration of classroom routines by expert teachers. *Curriculum inquiry, 17* (2), 135-176.

Lewis, A. C. (1995). An overview of the standards movement. *Phi Delta Kappan, 76*(10), 744-50.

Lefkowits, L. & Arens, S.A. (2004). *The national dialogue on standards-based education*: *Final report.* Aurora, CO: Mid-Continent Research for Education and Learning.

Li, J. (2007). Bridging across the mile-wide and mile-deep chasm: Living and coping with standards-based reform in Science education. In McInerney, D. M., Van Etten, S., & Dowson, M. *Standards in education*. Charlotte, NC: Information Age Publishing.

Luyten, H. (1994). Stability of school effects in Dutch secondary education: The impact of variance across subjects and years. *International Journal of Educational Research, 21*(2), 197-216

Luyten, J.W. & Snijders, T.A.B. (1996). School effects and teacher effects in Dutch elementary education. *Educational Research and Evaluation*, *2*: 1-24.

Magnusson, D., Bergman, L. R., Rudinger, G. & Torestad, B. (eds) (1991). *Problems and methods in longitudinal research: Stability and change*. Cambridge: Cambridge University Press.

Mansilla, P.U., Riejos, A.M.R. (2007). The European framework of languages: A piloting sample of cross-curricular strategy. *Higher Education in Europe, 32:2/3*, p 193 – 203.

Markowitsch, J., & Luomi-Messerer, K. (2008). Development and interpretation of descriptors of the european qualifications framework. *European Journal of Vocational Training, 42*(3-431), 33-58.

Marzano, R.J. (1996). *The rise and fall of standards-based education. Issue in brief.* Alexandria, VA: National Association of State Boards of Education.

Marzano, R.J. (1998). *Models of standards implementation*: *implications for the classroom*. Aurora, CO: Mid-Continent Regional Educational Lab.

Marzano, R.J. (1999). Building curriculum and assessment around standards. *The High School Magazine,* 6 (5), 324 – 326.

Marzano, R.J. (2000). *Transforming classroom grading.* Alexandria, VA: Association for Supervision and Curriculum Development.

Marzano, R. J., Kendall, J. S.(1996a). *A comprehensive guide to designing standards-based districts, schools, and classrooms.*Aurora, CO: Mid-Continent Regional Educational Lab. &Association for Supervision and Curriculum Development.

Marzano, R. J., Kendall, J. S.(1996b). *The fall and the rise of standards-based education. Issues in brief*. Aurora, CO: Mid-Continent Regional Educational Lab.

Marzano, R.J. (2007). *The art and science of teaching*. Alexandria, VA: Association for Supervision and Curriculum Development.

Mathison, S., & Freeman, M. (2008). T*eachers working with standards and state testing. In* Mathison, S., & Ross, E. W. *The nature and limits of standards-based reform and assessment: Defending public schools*. New York etc.: Teachers College, Columbia University.

Mathison, S., & Ross, E. W. (2008). *The nature and limits of standards-based reform and assessment: Defending public schools*. New York etc.: Teachers College, Columbia University.

Maulana, R., Opdenakker, M., den Brok, P., & Bosker, R. (2011). Teacher-student interpersonal relationships in Indonesia: Profiles and importance to student motivation. *Asia Pacific Journal of Education, 31*(1), 33-49.

McClure, P. (2005). Where standards come from. *Theory into Practice, 44*(1), 4-10.

Mohandas, R. (2004). *Overview of Education for Sustainable Development (ESD) in Indonesia*. Jakarta: National Office for Educational Research and development, The Ministry of National Education.

Money, S.M. (1992). *What is teaching effectiveness? A survey of student and teacher perceptions of teacher effectiveness*. Ontario: Humber College of Applied Arts and Technology.

Monk, D.H. (1994). Subject area preparation of secondary Mathematics and Science teachers and student achievement. *Economics of Education Review 13*(2): 125 – 145.

Muijs, D. (2006). Measuring teacher effectiveness: Some methodological reflections. *Educational Research and Evaluation. 12*(1), 53-74.

Muijs, D., Armstrong, P. & Chapman, C. (2010). *The impact of Teach First Teachers in Schools: Final Report*. London: The Maximum Impact Project for Teach First.

Muijs, D. & Reynolds, D. (2000). School effectiveness and teacher effectiveness in Mathematics: Some preliminary findings from the evaluation of the Mathematics Enhancement Program (Primary). *School Effectiveness and School Improvement, 11*(3), 273 – 303.

Muijs, D. & Reynolds, D. (2010). *Effective teaching: Research and practice.* London: Paul Chapman.

Muijs, D. & Reynolds, D. (3rd ed) (2011). *Effective teaching: Evidence and practice*. London: SAGE.

National Commission on Excellence in Education (1983). *A nation at risk: The imperative for educational reform.* Washington, D.C.: US Department of Education.

National Research Council. (1996). *National science education standards.* Washington, DC: National Academy Press.

National Research Council. (2001). *National science education standards*. Washington, DC: National Academy Press.

Needels M.C. & Gage, N.L. (1991). Essence and accident in process-product research on teaching. In Waxman, H.C. & Walberg, H.J. (eds). *Effective teaching: Current research.* Berkeley, CA: McCutchan Publishing Corporation.

Nelson, M. C., Cordray, D. S., Hulleman, C. S., Darrow, C. L., & Sommer, E. C. 2010, March). *A procedure for assessing fidelity of implementation in experiments testing educational interventions.* Paper presented at the annual meeting of the Society for Research on Educational Effectiveness, Washington, D.C.

Nielsen, H.D. (2003). Reforms to Teacher Education in Indonesia: does more mean better? In Beauchamp, E.R. (ed). *Comparative Education Reader.* New York: Routledge Falmer.

Neumann, K., Fischer, H. E., & Kauertz, A. (2010). From PISA to educational standards: The impact of large-scale assessments on science education in Germany. *International Journal of Science and Mathematics Education, 8*(3), 545-563.

Nurweni, A., & Read, J. (1999). The English vocabulary knowledge of Indonesian university students. *English for Specific Purposes, 18*(2), 161-175.

Organisation for Economic Co-operation and Development (OECD). (1995). *Performance standards in education: in search of quality*. Paris: OECD.

Paris, S.G. & Paris, A.H. (2001). Classroom applications of research on self-regulated learning. *Educational Psychologist, 36* (2): 89-101.

Padak, N. (2002). *Strategies that work: What does the evidence tell us? research to practice*. Columbus: Ohio State Literacy Resource Centre, Ohio State Department of Education.

Pemerintah RI. (2005). *Peraturan Pemerintah nomor 19 Tahun 2005 tentang standard nasional pendidikan*. Jakarta: Pemerintah RI.

Perrott, E. (1992). *Effective teaching*. New York: Longman Inc.

Pilot, Albert. (2007). *The Teacher as a Crucial Factor in Curriculum Innovation*: *the case of Utrecht University*. Paper presented at the Conference "Teaching and Learning according and after Bologna", at the Swiss Federal Institute of Technology (ETH, Zurich, Switzerland, 2007, March 9-10).

Pittelman, S.D. & Heimlich, J.E. (1991). Teaching vocabulary. In Hayes, B.L. (eds). *Effective Strategies for Teaching Reading*. Boston: Allyn & Bacon.

Plecki, M. (1991). *The relationship between elementary school size and student achievement*. Available at http://www.eric.ed.gov.proxy-ub.rug.nl/contentdelivery/servlet/ERICServlet?accno=ED396861.

Popham, W. J. (1997). The standards movement and the emperor's new clothes. *NASSP Bulletin, 81*(590), 21-25.

Powell, A. (1980). *The uncertain profession: Harvard and the search for educational authority*. Cambridge, Massachusetts: Harvard University Press.

Prasetiyo, B. (2009). Mengefektifkan kemampuan guru dalam menerapkan KTSP melalui MGMP Swadaya. *Jurnal Penddikan Inovatif,* volume 4 number 2, pp. 48-53.

Priyanto, A.D. (2009). Maximizing SFL Contriburion to ELT Indonesia. *UAD TEFL Conference*. Yogyakarta: UAD Press.

Proefriedt, W.A. (2008). *High expectations: The cultural roots of standards reform in American Education*. New York, NY: Teachers College Press.

Purkey, S., & Smith, M. (1983). Effective schools: a review. *The Elementary School Journal*, 83 (4), 427 – 452.

Raizen, S.A. (1998). Standards for science education. *Teachers College Record 100* (1), 66 – 121.

Rasbash, J., Charlton, C., Browne, W. J., Healy, M., & Cameron, B. (2005). *MLwiN Version 2.0.* Centre for Multilevel Modeling, University of Bristol.

Ravitch, D. (1995). *National standards in American education: A citizen's guide.* Washington, DC: Brookings Institution.

Reynolds, D. & Muijs, D. (1999). *Numeracy: an annotated bibliography for schools and colleges*. London: DfEE.

Reynolds, D., Sammons, P., De Fraine, B., Townsend, T., & Van Damme, J. (2011). *Educational Effectiveness Research (EER): A state of the art review*. Paper presented at the annual meeting of the International Congress for School Effectiveness and Improvement 2011 (Cyprus, 4-7 January, 2011). Available at http://www.icsei.net/icsei2011/State of the art/State of the art Session A.pdf.

Richman, S. (2001). *Parent power*: *why national standards won't improve education*. *Policy analysis no. 396*. Washington, DC.: Cato Institute.

Rosenshine, B. (1980). Skill hierarchies in reading comprehension. In Spiro, R. J., Bruce, B. C., Brewer, W.F (eds). *Theoritical Issues in Reading Comprehension; Perspectives from Cognitive Psychology, Lingustics, Artificial Intelligence, and Education*. New Jersey: Lawrence Erlbaum Associates, Publishers.

Rosenshine, B. (1983). Teaching functions in instructional programs. *Elementary School Journal 83* (4).

Rosenshine, B. & Stevens, R. (1986). Teachingfunctions. In M.C. Wittrock (ed.), *Handbook of research on teaching*, 3rd ed. Pp. 376-391. New York: Macmillan.

Rowe, K.J. (2007). *Educational effectiveness: The importance of evidence-based teaching practices for the provision of quality teaching and learning standards*. In McInerney, D. M., Van Etten, S., & Dowson, M. *Standards in education*. Charlotte, NC: Information Age Publishing.

Ruspini, E. (2002). *Introduction to longitudinal data*. London and New York: Routledge Taylor & Francis Group.

Sammons, P. & Hillman, J. & Mortimore, P. (1995). *Key Characteristics of Effective Schools: A Review of School Effectiveness Research*. London: Office for standards in Education (OFSTED).

Sandholtz, J.H., Ogawa, R.T., & Scribner, S.P. (2004). Standards gap: Unintended consequences of local standards-based reform. *Teachers College Record 106*, 1177 – 1202.

Scheerens, J. (1992). *Effective schooling: Research, theory, and practice*. London: Cassel.

Scheerens, J. & Bosker, R. J. (1997). *The foundations of educational effectiveness*. Oxford: Pergamon etc.

Scheerens, J. & Creemers, B.P.M. (1989). Conceptualizing school effectiveness. *International Journal of Educational Research, 13*, 691 – 706.

Schmidt, W. H., Houang, R., Shakrani, S. (2009). *International lessons about national standards*. Washington DC: Thomas B. Fordham Institute.

Schmoker, M. & Marzano, R.J. (1999). Realizing the promise of standards-based education. *Educational Leadership* (March).

Scriven, M. (1994). Duties of the teacher. *Journal of Personnel Evaluation in Education, 8*: 151-184.

Seidel, T., & Shavelson, R. J. (2007). Teaching effectiveness research in the past decade: The role of theory and research design in disentangling meta-analysis results. *Review of Educational Research, 77*(4), 454-499.

Setiogi, S.P. (2003, December 6). Final exam results show real educational quality. *The Jakarta Post*. Retrieved from https://www.thejakartapost.com/news/2003/06/12/final-exam-results-show-real-education-quality.html.

Shadish, W.R., Cook, T.D., & Campbell, D.T. (2002). *Experimental and quasi-experimental sedigns for generalized causal inference*. Boston, New York: Houghton Mifflin Company.

Shafiq, M. N. (2011). *Gender gaps in mathematics, science and reading achievements in muslim countries: Evidence from quantile regression analyses.* Education Economics, Forthcoming. Available at SSRN: http://ssrn.com/abstract=154782.

Slavin, R.E. (2010). Experimental studies in education. In Cremeers, B.P.M., Kyriakides, L., & Sammons, P. (eds). *Methodological advances in educational effectiveness research: quantitative methodology series*. Oxon: Routledge.

Smith, F., Hardman, F., Wall, K., & Mroz, M. (2004). Interactive whole class teaching in the national literacy and numeracy strategies. *British Educational Research Journal, 30*(30): 395-411.

Smith, L. and Glass, G.V. (1987). *Research and evaluation in education and the social sciences*. Englewood Cliffs, N.J.: Prentice-Hall.

Smith, L. & Land, M. (1981). Low-inference verbal behaviors related to teacher clarity, *Journal of Classroom Interaction 17*: 37-42.

Smith, L. & Sanders, K. (1981). The effects on student achievement and student perception of varying structure in social studies content. *Journal of Educational Research, 74*, 333-336.

Snijders, T. A. B., & Bosker, R. J. (1999). *Multilevel analysis: An introduction to basic and advance multilevel modeling.* London: Sage.

Snow-Renner, R. (2001). *Teachers' perspectives on standards-based education: Initial findings from a high-performing, high-needs school district.* Aurora, CO: Mid-Continent Research for Education and Learning.

Soar, R.S. & Soar, R.M. (1979). Emotional climate and management. In Peterson, P & Walbeg, H. (eds), *Research on teaching concept: Findings and implications*. Berkeley, CA: McCutchan.

Spillane, James P., 2004. *Standards Deviation; how school misunderstood education policy*. Harvard University Press.

Stalling, J., Corry R., Fairweather, J., & Needels, M. (1978). *A study of basic reading skills taught in secondary school*. Menlo Park: SRI International.

Suara Merdeka (2009, September 16). Belum semua sekolah terapkan KTSP. Retrieved fromhttp://suaramerdeka.com/v1/index.php/read/cetak/2009/09/16/80739/Belum.Semua.Sekolah.Terapkan.KTSP on November 13, 2009.

Subotnik, R.F., & Walberg, H.J. (2006). *The scientific basis of educational productivity.* Greenwich, CT: Information Age Publishing.

Suharto (2012, August 9). *Refleksi 5 tahun program sertifikasi guru*. Retrieved online from http://www.m-edukasi.web.id/2012/08/refleksi-5-tahun-program-sertifikasi.html.

Superfine, B.M. (2005). The politics of accountability: The rise and fall of goals 2000. *American Journal of Education 112* (November 2005), pp. 10 – 43.

Sulistiani, E. (2009). *The Perception of English Teachers at Public JuniorHigh Schools in Malang on the Implementation of School-Based Curriculum(SBC).* Unpublished thesis, English Language Education, Faculty of Letters, State Universityof Malang.

Sumintono, B. (2006). *Decentralized Centralism: School Based Management Policies and Practices at State Secondary Schools in Mataram, Lombok, Indonesia*. Unpublished doctoral dissertation. New Zealand: Faculty of Education at the Victoria University of Wellington.

Superfine, B.M. (2005). The politics of accountability: the rise and fall of goals 2000. *American Journal of Education*. Vlume 112 No. 1, p. 10 – 43.

Teddlie, C. & Reynolds, D. (2000). *The international handbook of school effectiveness research*. London: Falmer Press.

Teddlie, C. & Stringfield, S. (1993). *School make a difference: Lesson learned from a 10-year study of school effects.* New York: Teachers College Press.

Tienken, C.H. (2011). Common Core State Standards: An Example of Data-Less Decision Making.AASA Journal of Scholarship & Practice, 7 (4), 3-18.

Tilaar, H.A.R. (1992). *Manajemen Pendidikan Nasional, kajian pendidikan masa depan (managing national education, a study of future education)*. Bandung: Remaja Rosdakarya.

Thair, M. & Treagust, D.F. (2003). A brief history of a science teacher professional development initiative in Indonesia and the implications for centralized teacher development. *International Journal of Educational Development 23 (2003) 201-213.*

Thompson, C. J. (2009). Preparation, practice, and performance: An empirical examination of the impact of standards-based instruction on secondary students' math and science achievement. *Research in Education, 81*(1), 53-62.

Thompson, S. (2001, January). The authentic standards movement and its evil twin. *Phi Delta Kappan*, 358-362.

Tuhusetya, S. (2007). *Pekik merdeka dan KTSP*. Retrieved online from http://sawali.info/2007/08/02/pekik-%E2%80%9Cmerdeka%E2%80%9D-di-tengah-perubahan-paradigma-pendidikan/.

US Department of Education. (2008). *A nation accountable: Twenty-five years after "A nation at risk".* Washington D.C.: US Department of Education.

Utomo, E. (2005). *Challenges of Curriculum Reform in the Context of Decentralization: The Response of Teachers to Competency Based Curriculum (CBC) and Its Implementation in Schools*, unpublished doctoral dissertation, University of Pittsburgh .

Van Damme, J., Opdenakker, M.C., Van Landeghem, De Fraine, Pustjens, & Van de Gaer. (2006). *Educational effectiveness: Introduction to international and Flemish research on schools, teachers, and classes*. K.U. Leuven: Centre for Educational Effectiveness and Evaluation. Retrived from https://lirias.kuleuven.be/bitstream/123456789/156919/2/ChaireFrancqui2006EducationalEffectiveness.pdf.

Van der Werf, M.P.C. (1995). *The Educational Priority in the Netherlands: content, implementation and outcomes.*The Hague: SVO.

Van der Werf, M.P.C., Creemers, B.P.M., De Jong, R., & Klaver, L. (2000). Evaluation of school improvement through an educational effectiveness model: The case of Indonesia's PEQIP project. *Comparative Education Review, 44*(3), p. 329–356.

Van der Werf, M.P.C., Opdenakker, M-C., & Kuyper, H. (2008). Testing a dynamic model of student and school effectiveness with a multivariate multilevel latent growth curve approach. *School Effectiveness and School Improvement, 19*(4), 447-462.

Wanacott, M.E. (2000). *Standards: An Embarrassment of Riches. In brief: Fast facts for policy and practice No. 4*. Columbus, OH: National Dissemination Center for Career and Technical Education.

Wang, M. & Walberg, H. (1991). Teaching and educational effectiveness: Research synthesis and consensus from the field. In K. J. Rehage (Series Ed.) & H. C. Waxman & H. J. Walberg (Vol. Ed.), Effective teaching: Current research (pp.63-80). Berkeley, CA: McCutchan Publishing.

Wehby, J. H., Maggin, D. M., Johnson, L., & Symons, F. J. (2010). *Improving intervention implementation and fidelity in evidence-based practice: Integrating teacher preference into intervention selection.* Evanston, IL: Society for Research on Educational Effectiveness.

Wheelock, A. (1995). "*Standard-Based Reform:* w*hat does it mean for the middle grades*? New York: Edna MsConnell Clard Foundation.

Widmer, T. (2004). The development and status of evaluation standards in western Europe. *New Directions for Evaluation,* (104), 31-42.

Wigfield, A., &. Guthrie, J.T. (1997). Relations of children's motivation for reading to the amount and breadth of their reading. *Journal of Educational Psychology 89*: 420–32.

Winarsih (2008). *Implementasi kebijakan sertifikasi guru sekolah dasar (studi kasus di Kabupaten Semarang*). Unpublished master thesis. Semarang: Program Pasca Sarjana Universitas Diponegoro. Retrieved from http://eprints.undip.ac.id/17753/1/WINARSIH.pdf.

Wise, A. E.& Darling-Hammond, L (1983). *Beyond standardization: State standards and school improvement*.Washington, DC: Dingle Associates, Inc.

Wong, R. M. H. (2007). Motivation and English attainment: A comparative study of Hong Kong students with different cultural backgrounds.The Asia-Pacific Education Researcher, *16*(1), 45-60.

Wragg, E.C. (1984) (ed.) *Classroom teaching skills.* London: Croom Helm.

Wurtz, E., & Others (1993). *Promises to keep: Creating high standards for American students. Report on the review of education standards from the goals 3 and 4 technical planning group to the national education goals* panel. Washington DC: National Education, G. P.

Zuzovsky, R., & Libman, Z. (2006). Standards of teaching and teaching tests: Is this the right way to go? *Studies in Educational Evaluation, 32*(1), 37-52.

APPENDICES FOR CHAPTER 3

ELABORATED STANDARDS OF CONTENT READING COMPREHENSION IN ENGLISH FOR JUNIOR SECONDARY SCHOOL

Siti Nurul Azkiyah

FACULTY OF TEACHER TRAINING AND EDUCATION, UIN JAKARTA
INSTITUTE FOR EDUCATIONAL RESEARCH (GION), RUG, NL
July 2010

Table of Content

INTRODUCTION

As a mandate of the new law No. 20/2003 on National Standards of Education and in line with the standards movement in other countries, the Indonesian government has set up eight National Standards of Education in 2005, in order to improve the quality of education. Two of these are the standards of graduate competence and the standards of content, which should guide teaching and learning processes including assessment. The new policy requires schools and teachers to develop their own curriculum, syllabus, and lesson plan based on these two standards. Many teachers have been reported to have problems dealing with both the development and the implementation of the curriculum due to lack understanding of the concept, minimum information dissemination as well as not enough supporting resources (Hanafie, 2007, Lampung post, 2006, Prasetiyo, 2009, Suara Merdeka, 2009, Sulistiyani, 2009).

This implies that teacher professional development program is indeed necessary. However, many teachers do not have access to such programs due to various reasons such as less quota of participation, financial problem, as well as geographical area (Cahyono, 2008, Jazadi, 2003). Thus there is a gap between the need to improve the quality of teachers and the available and accessible programs. Therefore, this document argues that one of the basic and fundamental improvements needed is providing clear, specific, and measurable standards of graduate competence and content to guide teachers' planning and delivery as well as assessment of their teaching. Moreover, some problems are found in the government standards document. For instance the two standards are dominated by the terms understanding the meaning of and responding to meaning without further explanation on what they actually mean. In this document they are made more specific, concrete and measurable.

Since the national exam measures only reading comprehension, the document is limited to reading in English in Junior Secondary School (JSS) grade VIII. However, the document is accessible for the other two grades (grade VII and IX). It is expected that this document will give clearer ideas on what to plan, to teach, and to assess when teaching reading comprehension.

This document will present three main sections:

1. The government standards of graduate competence and content of English especially on reading in Junior Secondary School;
2. Analysis of the government standards;
3. The elaboration of the standards of content.

An appendix on the examples of text and questions representing the elaborated standards is provided. In addition, it should be noted that the elaborated standards in this document are the minimum requirement that all children should be able to do. It is true that students have different personal and academic background, which may affect them in achieving the standards. However, the standards movement requires teachers to find various strategies that enable them to help different students to achieve the goals.

1. The Government Standards of Garduate Competence (SKL) and Content (SI)

The SKL mandates that students are able to perform in four language skills, which are listening, speaking, reading, and writing. With respect to reading as the focus of the document, the SKL is:

"understanding meaning in simple written interpersonal and transactional discourses both formal and informal in the form of recount, narrative, procedure, descriptive, and report, in daily life context."

They are made more concrete in SI, which is divided into the standards of competencies and basic competencies. With respect to reading, the standards are as follow:

Table 1 Standards of Content of English (reading) for JSS

Grade/ Semester	Standards of Competence	Basic Competencies
VII/1	Understanding the meaning of short and very simple functional written texts related to the closest environment.	1. **Meaningful reading aloud words, phrases, and sentences related to the closest environment with correct pronunciation, stress and intonation.** 2. **Responding to meaning** to short and very simple functional texts related to the closest environment accurately, fluently, and meaningfully.
VII/2	Understanding the meaning of short and very simple functional written texts and essay in the form of descriptive and procedure related to the closest environment.	1. **Responding to meaning** in short and simple functional texts related to the closest environment accurately, fluently, and meaningfully. 2. **Responding to meaning and rhetorical stages** to short and very simple descriptive and procedural essays related to the closest environment. 3. **Meaningful reading aloud** short and very simple descriptive and procedural essays related to the surroundings with correct pronunciation, stress and intonation

Grade/ Semester	Standards of Competence	Basic Competencies
VIII/1	**Understanding the meaning** of short and simple functional written texts in the form of descriptive and recount related to the surroundings.	1. **Meaningful reading aloud** short and simple functional texts and essays in the form of descriptive and recount related to the surroundings with correct pronunciation, stress and intonation. 2. **Responding to meaning** to short and simple functional texts related to the surroundings accurately, fluently, and meaningfully 3. **Responding to meaning** and rhetorical stages to short and simple essays in the form of descriptive and recount related to the surroundings
VIII/2	**Understanding the meaning** of short and simple essays in the form of descriptive and recount in order to be able to interact with the surroundings.	1. **Meaningful reading aloud** short and simple functional texts and essays in the form of recount and narrative related to the surroundings with correct pronunciation, stress and intonation. 2. **Responding to meaning** to short and simple functional texts related to the surroundings accurately, fluently, and meaningfully. 3. **Responding to meaning** and rhetorical stages to short and simple essays in the form of narrative and recount related to the surroundings.
IX/1	**Understanding the meaning** of simple and short functional texts essays in the form of procedure and report related to daily life context.	1. **Meaningful reading aloud** short and simple functional texts and essays in the form of procedure and report related to the daily life context with correct pronunciation, stress and intonation. 2. **Responding to meaning** to short and simple functional texts related to the daily life context accurately, fluently and meaningfully. 3. **Responding to meaning** and rhetorical stages to short and simple essays in the form of procedure and report related to the daily life context.

Grade/ Semester	Standards of Competence	Basic Competencies
IX/2	**Understanding the meaning** of simple and short functional texts and essays in the form of narrative and report related to the daily life context.	1. **Meaningful reading aloud** short and simple functional texts and essays in the form of narrative and report related to the daily life context with correct pronunciation, stress and intonation 2. **Responding to meaning** to short and simple functional texts related to the daily life context accurately, fluently, and meaningfully 3. **Responding to meaning** and rhetorical stages to short and simple essays in the form of narrative and report related to the daily life context

2. Analysis of the Government Standards

First of all, the SKL does not specify the competencies for each grade. This could be due to the reason that the competencies are to achieve when students are about to graduate. However, the regulation should mention it. Secondly, they are general and broad. For example they do not specify how many words students in each level should master. In addition, in reading for instance, they do not specifically explain what reading skills students are expected to perform (for instance to get word meanings across the text, to get the main idea, to infer something from the text etc.). The standards of content are actually expected to give more concrete concept. Yet, as seen in table 1, the terms used are *understanding the meaning of* and *responding to meaning* without any further explanation on what they exactly mean.

However, the standards of content indeed provide the kind of text and theme students are expected to learn in each level. The differences among grades are the length (very short and short), the difficulty level (simple and very simple), the type of the text and the theme (the closest environment, the surroundings and the daily life context). In this case, the terms need to be clarified are *short, simple, and very simple, the closest environment*, *the surrounding*, and *the daily life context*.

Another crucial problem is the inconsistency of the terms used. In the SKL and basic competencies of SI, the competence is limited to understanding meaning, whereas the basic competencies of SI reach responding to meaning which is higher than understanding meaning. It is logical to assume that a student will be able to respond when he/she understands a meaning. The standards of competence for grade VII semester 1, for example, is limited to "understanding the meaning of short and very simple functional written texts related to the closest environment". However, the basic competence, which is supposed to be lower in terms of the policy level, mention not only understanding meaning but also responding to meaning.

So far, there has not been any study dealing with the clarification of these terms. Therefore, the paper advocates the importance of elaborating the standards into a clear and measurable document and at the same time also offers the elaborated standards. The elaboration attempts to clarify the aforementioned terms so that teacher will have clearer ideas on what skills to be taught and assessed when delivering their teaching.

3. The Elaboration of the Standards of Content

3.1 Description of Understanding Meaning and Responding to Meaning

This section will clarify what understanding and responding to meaning exactly means in real teaching, learning, and assessment. Literature study on reading comprehension and analysis of competence taught in three books used by teachers in the sample schools (Madrasah Tsanawiyah/MTs in Jakarta and Banten Provinces) and those assessed in the national exam are carried out. Rosenshine (1980) provides observable skills of reading comprehension, which falls into three general types from easier to more complex skills. This categorization is useful to understand which specific reading skills belong to "understanding meaning" and "responding to meaning". "Understanding meaning" consists of easier reading skills while "responding to meaning" consists of more complex skills, as shown in the following table.

Table 2 Rosenshine's and the Indonesian standards' categorization of reading skills

No.	Rosenshine's Categorization	The Indonesian Standards' Categorization
1.	Locating details: the simplest skill, which involve recognition, paraphrase, and/or matching	**Understanding meaning**
2.	This skill can include answering specific text-based questions, reading tables/number/pictures, identify and describe characters, identify types and pattern of text structure/development. Simple inferential skills: understanding words in context	
3.	Simple inferential skills: inferring sequences, which can include: a. Recognizing the sequence of events b. Recognizing cause and effects relationships c. Comparison and contrasting	Responding to meaning
4.	Complex Inferential Skills a. Recognizing the main idea/title/topic b. Drawing conclusions c. Predicting outcomes	

It should be noted that Rosenshine's original categorization consists of only three categorizations (locating details, simple inferential skills and complex inferential skills) since the simple inferential skills are grouped into one category. In this document, understanding words in context is considered as "understanding meaning" and therefore it is separated from the other simple inferential skills mentioned by Rosenshine in order to make it easier to identify the skills under "understanding meaning" and those under "responding to meaning".

The analysis of competencies taught in Indonesian textbooks and assessed in the national exam also reflects Rosenshine's classification of reading skills. The following table describes reading skills taught in the Indonesian textbooks and its category according to Rosenshine as stated in table 1.

Table 3 Reading skills taught in Indonesian textbooks and their categorization according to Rosenshine

No.	Description of skills (based the analysis of Indonesian book)	Grade VIII	Grade IX	Category/Sub-skills
1.	Reading aloud with correct stress and intonation	X	X	Non-reading comprehension skill
2.	Finding topic / general ideas of a text or a paragraph and supporting ideas	X	X	4a. Recognizing the main idea/title/topic
3.	Predicting title	X	X	4c. Predicting outcomes
4.	Identifying tenses and other grammar related aspects used in the text and the reasons of using them	X		Non-reading skills
5.	Identifying types and stages of text development (setting, list of events, reorientation etc)	X		1. Location details/ identifying types and pattern of text structure
6.	Identifying the goals of the text	X		4b. Drawing conclusion on the purpose of the author)
7.	Identifying and describing characters	X		1. Locatiting details/ answering specific text-based question
8.	Answering text-based questions (4 W 1 H - who, what, when, where, how) or factual information	X	X	1. Locating details/ answering specific text-based question
9.	Identifying/interpreting writer's opinion about people/events described in the text	X		4b.Drawing conclusion/ identifying the weiter's opinion/attitude

10.	Confirming that statements are true or false according to the text	X		4b. Drawing conclusion/ evaluating ideas presented in the text
11.	Matching pictures with the right description or vice versa	X	X	1. Locating details/ reading pictures
12.	Identifying meaning of certain words in a text	X	X	2. Undestending words in context
13.	Reflecting ideas in the text (whether we can do what is suggested in the text		X	4b. Drawing conclusion
14.	Giving opinion about the text		X	4b. Drawing conclusion
15.	Identifying information in chart, diagram, table, map		X	1. Locating details/ reading table or number
16.	Identifying words and phrases related to the theme		X	4a. Recognizing the main idea/title/ topic

Table 4 Competencies assessed in the national exam and their categorization according to Rosenshine

No	Description of Skill	02/03	03/04	04/05	06/07	07/08	Category/Sub-skills
1.	Answering text-based questions (WH - who, what, when, where, how) or factual information	X	X	X	X	X	1. Locating details/ answering specific text-based question
2	Finding pronounce reference	X	X	X		X	2. Understanding words in context
3.	Responding given situation in a dialog						4c. Predicting outcomes
	a. Apologize for mistake	X					
	b. Asking for /offering help	X		X	X		
	c. Agreement/disagreement	X	X	X			
	d. Giving Appropriate instruction		X				
	e. Showing admiration		X				
	f. Introduction			X			
4.	Grammar related aspect						Non-reading skills
	a. Singular vs plural	X	X	X			
	b. Tenses	X	X	X	X		
	c. Comparative degree	X	X	X	X		
	d. Conjunction	X			X		

No	Description of Skill	02/03	03/04	04/05	06/07	07/08	Resenshine
5	Drawing conclusion a. From text A, it can be concluded that B is	X				X	4b. Drawing conclusion
	b. Certain type of ads should be placed in certain place	X					
	c. Conclusions from simple statistical numbers presented in a table/matrix or pictures (price, number of people, ages)	X		X	X	X	
	d. Judgment that things are nice, bad, cheap, expensive and the degree of certainty - rather nice, probably etc	X		X			
6.	Identifying word / phrase meaning in the text	X	X	X	X	X	2. Understanding words in context
7.	Finding theme/topic of a text and a paragraph	X	X	X	X	X	4a. Ricognizing the main ideas/title/ topic
8.	Predicting title	X	X	X	X	X	4c. Prdicting outcomes
9.	Identifying the mood of the writer	X					4b. Drawing conclusion
10.	Evaluating statements (true/false) according to the text	X	X			X	4b. Drawing conclusion
11.	Matching pictures with the right description or vice versa		X	X		X	1. Locating details/ reading picture
12.	Arranging sentences into a paragraph		X	X	X	X	1. Locating details/ identify types and text development
13.	Finding moral values in the text					X	4b. Drawing conclusion

Few of the listed reading skills in the two tables cannot be regarded as reading skills. For instance grammar related aspects in point 3 in table 3 and point 4 in table 4. They should be regarded as prerequisite to comprehension. It is true that grammatical structure will affect comprehension and therefore it is important to teach grammar. However, it is important to note that comprehension questions should go beyond grammatical questions. Similarly, reading aloud with correct stress and intonation (point 1, table 3) is not reading comprehension question either. It is also a precondition for reading comprehension. Lastly, arranging sentences into a paragraph was possibly meant to be a writing test. Yet, writing is not simply arranging sentences into a paragraph. In this case it is assumed that students will be able to perform such task when they understand the stages of text development, which is close to locating details especially identifying the structure of text development.

More important is the fact that the skills taught in Indonesian textbooks and assessed in the national exam are dominated by locating details and drawing conclusions. Other skills are not equally taught and assessed. Therefore, it is suggested that the distribution of skills taught should be balanced among different skills. Moreover, one skill is not found at all in both textbooks and national exam, which are simple inferential skill especially inferring sequences.

Now we come to the main issue of this document. Referring to the work of Rosenshine and the analysis of reading skills taught in Indonesian books and assessed in the national exam, what are specific and measurable competencies in the government standards? They are presented in the following table. The numbering of the category described in the first column does not follow the number of Rosenshine's categorization (table 2) and thus new numbering is used.

Table 5 Specific competencies for understanding and responding meaning and their assessment checklist

Category	Specific competence	Assessment checklist
	Understanding Meaning (UM)	
1. Understanding words in context.	1. Recognizing words/ paraphrase meanings in the text	• Identify the meaning of most (80%) words / phrases in the text • Identify pronounce reference
2. Locating details	2a. Reading pictures, tables, numbers 2b. Answering specific text-based questions 2c. Identify and describing characters 2d. Identify types and pattern of text structure such as: - Sequences of events for narrative, procedure, recount - Description	• Matching pictures with the right description or vice versa • Identifying the meaning of pictures and signs • Identifying information in a chart, diagram, table, map • Locating information from a text into a table or vice-versa • Identifying factual information explicitly stated in the text. Usually they include 5 W 1 H questions (who, what, when, where, why, how) • Identifying all persons/ characters mentioned in the text • Describing characters • Identifying key words that signal text structure • Reorganizing scrambled sentences into a paragraph or scrambled paragraphs into a text

Responding to Meaning (RM)		
3. Recognizing the main idea/ topic/supporting ideas	3. Finding main ideas/ topic/supporting details	• Identifying words and phrases related to the theme • Find topic / general idea of a text or paragraph and supporting details
4. Drawing conclusion	4a. Finding author's purposes, attitude, tone, or mood	• Identify the goals of the text • Identify the mood of the writer
	4b. Making inferences	• Interpreting the printed information in the text. For instance a text indicates that a character is preparing a swimsuit and swim glasses. We can conclude that the character is going to swim
	4c. Inferring sequences, which can include: • sequences of events • causal relationship • comparison and contrast	• Identifying action/incident/ event that may take place between two explicitly stated actions or events and identify what would happen next if the story did not end as it did (for sequences of events) • Identifying unstated cause and effect relationship (for causal relationship) • Identifying likeness and differences in characters, time, or places (for comparison and contrast)
	4d. Evaluating ideas presented in the text	• Confirming whether certain statements are true or false according to the text • Reflect ideas presented in the text (whether we agree or can do what is suggested in the text)
5. Predicting outcomes	5. Predict title and what comes or happens next in the text	• Predicting title • Predicting what the next paragraph will be about after reading or discussing previous paragraph • Predicting responses of given situation

Some of these reading skills may have been taught to students. However, it is necessary for a policy document such as the standards to be clear and specific so that teachers can figure out observable and measurable competencies to teach to students when they read the document. Moreover, not all books provide good guidance and materials for teaching. Priyanto (2009) on his critical review of two English course books, for instance, concludes that the books did not give clear point on the fact that language is the realization of a social process that always has a social purpose. Instead, they tend to focus on form, without giving any further explanation on why and for what purposes learners should use those forms. Another point he made was that grammar is presented separately, not to support the learners to achieve the social goals of using the language. However, he acknowledges that the books have presented the learners a set of well sequenced activities.

Lastly, despite the inconsistency on the use of the terms understanding and responding to meaning in the SKL and SI, this document considers that both understanding and responding to meaning are required in all grades. However, the standards differentiate the length and the difficulty level, the type and the theme of the text across grades, which should be used to scaffold the materials for each grade. They will be explained in the following subsection.

3.2 Length and difficulty level of text

It is true that length, type, and theme are some of the determinant factors to scaffold materials. They are interconnected to what experts analyze about the factors to consider when scaffolding materials such as vocabularies, sentence structure, length, coherence, familiarity of background knowledge required, and text structure.

With respect to the length and difficulty level as indicated by the simplicity of the text, the standards differentiate that short and very simple texts should be used in grade VII, whereas short and simple texts should be used in grade VIII and IX. What is the difference between very simple and simple texts in each grade? Of course texts in grade VII should be the easiest and shortest and gradually move to more complex and longer texts. In order to provide clues for teachers, this document considers the importance of building consensus by looking at the length of texts in each grade in various books. The texts for each grade vary across different books. In grade VII, the texts range from 37 up to 103 words, whereas in grade VIII they range from 92 words until 237 words. Students in grade IX usually read texts ranging from 92 – 292 words. In terms of the length or the number of words, this document concludes that the text should be not more than 150 words for grade VII, not more than 250 words for grade VIII and not more than 350 words for grade IX.

Next, the difficulty level of the text could be seen from the grammatical structure and the familiarity of vocabularies in the text. A text could be short but complicated especially when the structure is not simple and the vocabularies used are not familiar to students. Sentence structure deals with grammatical structure used in the text. There are four basic types of sentences:

1. Simple sentence : subject + verb + complete thought (Independent clause/IC)
   a. Some students like to study in the morning
   b. Rita and Adi do their homework after school
2. Compound sentences: IC + IC
   a. Rendy likes apples and Mila likes oranges.
   b. Ali goes to school and Mom goes to her office.
3. Complex sentence: Dependent clause (DC) + IC or vise versa. . A complex sentence always has a subordinator such as because, since, after, although, or when or a relative pronoun such as that, who, or which.
   a. The teacher returned the homework after she noticed the error.
   b. The students are studying hard because tomorrow they will have exam.
4. Complex-compound sentence: DC + IC + IC. It combines elements of compound and complex sentences, which usually is comprised of at least one dependent clause and two independent clauses.
   a. Although Diah prefers tea, she bought coffee and she drank it.
   b. Samsul forgets to return books to the library, so he rushed to go to the library when he finally remembered.

The Indonesian standards limit very simple texts for grade VII and simple texts for both grade VIII and IX. It is suggested that texts with mostly simple sentence structure is used in grade VII and moves to more complex sentence structure for higher grades.

Concerning the familiarity of the text, the Indonesian standards limit the theme of the text in a gradual step in accordance with students' interaction with their environment, which will be explained in the following section.

3.3 Type of text

The standards of content mention that the texts in all grades are functional text, which in a simple way can be defined as a text useful to find information used in our daily life. For instance if we want to make cookies, then we should read recipe or if we want to know our friend's number then we should read phone book and so on. Furthermore, they also have specific text for each grade.

The texts are limited to descriptive and procedural texts in grade VII, descriptive, recount, and narrative in grade VIII, and procedural, narrative, report in grade IX. When the types of the text are used in more than one grade such as narrative and procedural texts, other limitation such as length and difficulty level of text as well as theme should be used to scaffold the material. In this elaborated standards, the limitation used is the more specific type of text such as descriptive, narrative, procedural, report and the like.

3.4 Theme

As previously mentioned, the theme is limited to the closest environment for grade VII, the surroundings for grade VIII and the daily life context for grade IX. This limitation follows

a logical approach in which the closest environment is expected to be the most restricted and then it goes to broader context. However, it will be clearer for teachers when examples or further explanation on the limitation are provided.

Again, building consensus is carried out in order to understand the theme of the texts in each grade in the three analyzed books. The following table lists types, topics and specific competencies for each grade.

Table 6 Specific type and topic of text and competencies emphasized in each grade and semester

Grade/Theme	Type of Text	Specific Topic	Specific Competencies Emphasized
VII/Semester 1 The closest environment		• My identity • Introduction • My family • My house	**Understanding meaning** • Recognizing words/paraphrase meaning in the text • Answering specific text based questions • Reading simple pictures, tables **Responding to meaning** • Recognizing the main idea/ topic/upporting ideas
VII/Semester 2 The closest environment	• Descriptive • Procedure	• My school • My hobby • My favorites (food, places, persons)	**Understanding meaning** • Recognizing words/paraphrase meaning in the text • Answering specific text based questions • Identifying and describing characters • Identifying type and pattern of text development **Responding to meaning** • Recognizing the main idea/ topic/supporting ideas • Finding author's purposes, attitude, tone, or mood
VIII/Semester 1 The surroundings	• Descriptive • Recount	• Friendship • Travelling • Teenager's life	**Understanding meaning** • Recognizing words / paraphrase meaning in the text • Answering specific text based questions • Reading pictures, tables, number

Grade/Theme	Type of Text	Specific Topic	Specific Competencies Emphasized
			• Identifying types and pattern of text development **Responding to meaning** • Recognizing the main idea/ topic/supporting ideas • Finding author's purposes, attitude, tone, or mood making inferences • Inferring sequences • Predicting outcomes
VIII/Semester 2 The surroundings	• Recount • Narrative	• My experience • Profession • Shopping	**Understanding meaning** • Recognizing words / paraphrase meaning in the text • Answering specific text based questions • Identifying and describing characters, passage organization **Responding to meaning** • Recognizing the main idea/ topic/ supporting ideas • Finding author's purposes, attitude, tone, or mood • Evaluating ideas presented in the text • Predicting outcomes
IX/Semester 1 The daily life context	• Procedure • Report	• Environment • Technology • Arts	**Understanding meaning** • Recognizing words / paraphrase meaning in the text • Answering specific text based questions • Identifying and describing characters • Reading pictures, tables, numbers • Identify types and pattern of text development

Grade/Theme	Type of Text	Specific Topic	Specific Competencies Emphasized
			Responding to meaning • Recognizing the main idea/ topic/ supporting ideas • Finding author's purposes, attitude, tone, or mood • Evaluating ideas presented in the tex
IX/Semester 2 The daily life context	• Narrative • Report	• Sports • Mass media • Public services	**Understanding meaning** • Recognizing words / paraphrase meaning in the text • Answering specific text based questions • Identifying and describing characters • Reading pictures, tables, numbers **Responding to meaning** • Recognizing the main idea/ topic/ supporting ideas • Finding author's purposes, attitude, tone, or mood • Evaluating ideas presented in the text • Inferring sequences • Predicting outcomes

The listed topics in this table are to give ideas about the kind of topics teachers could introduce to students in each grade. It is, or course, possible for teachers to teach different topics. This document considers that the most important factor is the skills or the specific competencies that should be taught to students. The column on specific reading skills or competencies that could be emphasized in each grade and semester could be used to plan the materials and the delivery of teaching in the classroom. It is clear that there is no much different of skills emphasized across grades except grade VII in which the skills are dominated by those under "understanding meaning". This means that the difficulty level and the length of the text should be emphasized so that they will differentiate the three grades.

APPENDIX

EXAMPLE OF TEXT AND QUESTION

In order to make the discussion on the proposed elaborated standards more comprehensive and practical, examples of texts and questions are provided. They are taken from different books that some teachers in Jakarta and Banten use and therefore it is expected that they are not totally new for teachers. The questions are also taken from the books but added with brief analysis with respect to the proposed elaborated standards.

Text 1 (The Bridge English Competence Grade VII, p. 21)

This text is for grade VII semester 1.

Here is Tuesday's timetable of Nicky's class. Read it carefully to complete the timetable below. Then compare it with your friend's work.

Nicky's class has seven periods of lessons on Tuesdays. The first two are for English. Then, one period is for Geography continued by Mathematics for two hours separated by break time. The last two periods are for Civics.

Hour	Subject
07.00	
07.45	
08.30	
09.15	
BREAK	
10.15	
11.00	
11.45	

This text is very short as it consists of only 37 words and simple as most sentences use the formula of simple sentences. Therefore the text is used in grade VII semester 1. The question accompanying this text requires students to locate narrative information into a table. This is in accordance with specific competence no 4 of understanding meaning, which in this case is reading table and numbers (see table 4, p. 7). This exercise should not take long time as it is very simple and short. Teachers could ask students to do more similar activities to ensure that they achieve the intended skill.

Text 2 (English in Focus G. VII p. 30)

FOR SALE
Two storey houses. It is located near Patrol highway, JL. Sastrawan No. 3 Flores. The house consists of the following:

- A living room, a drawing room, a dining room, a study room, a kitchen, two bathrooms and a garage.
- Children's and nanny's bedrooms upstairs and a main bedroom downstairs.
- A sofa and some chairs in the drawing room.
- A refrigerator, an electric fan, and a television.
- Kitchen utensils, such as pots, pans, kettles, gas stoves and the kitchen sink

It also has 3000 V power, telephone line, hotspot, and water heater. Serious buyer contacts Tia 987654 (101 words).

Questions:

1. Could you mention the rooms in the house?
2. Where are the children's bedrooms?
3. What is there in the drawing room?
4. Can you mention some kitchen utensils?

This text is also aimed at students in grade VII, consisting of 103 words. The questions in this text are mainly **specific text-based questions** which answers are easily found in the text. Other questions could be explored, for instance places where this kind of advertisement could be found, the purpose of the writer (finding author's purpose) as well as possible writer of this advertisement (identifying characters). Teacher can also explore more questions which belong to the higher level (responding meaning) such as making inferences or prediction of possible reasons of the owner to sell her/his house.

Text 3 (The Bridge English Competence, VIII, p. 33)

There are several ways you can find out about the countries and places you want to visit. You can talk to your friends who have traveled to the places, or you can go and see film about them. You can also read travel books.

There are three kinds of travel books. The first are those that give a personal account of travels. If you are informative and have good index, they can be useful to you when you are planning your travels. The second kinds are those books whose purpose is to give an objective description of things to be done and seen. It can be classified as selective guidebook. The third kind is those books which are called 'a guide' to some place or other. If they are good, they will give an analysis or an interpretation. Like the first kind they are inspiring and entertaining. But their primary function is to assist the reader who wishes to plan in the most practical way.

Whatever kind of travel book you choose, you must make sure that it does not describe everything as magical. You must also note its date of publication. Affair and many things change quickly. Finally, you should make that the contents are well presented and easy to find. (212 words)

Questions:

1. What is the best title of the above text? (RM predicting title)
2. How many kinds of travel books are mentioned? What are they? (UM answering specific text-based Q)
3. What can we do if we wish to go to other countries? (RM making inferences)
4. Why must we note the date of publication? (RM making inferences)
5. How should the contents be? (UM answering specific text-based Q)
6. What kind of travel books is the best? (RM drawing conclusion)
7. What is the purpose of the second kind of the travel book? (UM answering specific text-based Q)
8. What does the word "you" refer to? (RM making inferences)
9. What does the word "it" refers to in paragraph 3? (RM making inferences)

The competencies that could be exercised using this text also vary. As can be seen in the questions, they range from recognizing words meaning up to more sophisticated skills such as predicting title, making inferences and drawing conclusion. This also indicate that more complicated text allow teachers to exercise more competencies.

Text 4 (English in Focus for Grade VIII, p. 135 - 136)

The Chalip and the Clown

The Chaliph of Baghdad hired an intelligent and high spirited man as his court jester, and was much amused by his clever, witty comments. The clown was so well loved by his master that everyone at court showed him great resepect. And so the Caliph was much surprised one day when he heard his beloved clown crying out in distress from the throne room. The Caliph hurried there and was astonished to find the guards beating the clown badly.

"Leave him alone at once"! he ordered. "Why are you beating him?"

"We found him sitting on your throne, O majesty!" the captain of the guards explained.

"Out of my sight!" commanded the Caliph. "For sure the jester did not do it with the intention of offending me."

The clown, however, continued to weep and wail even after the guards had left.

"Stop it! Said the irritated Caliph.

"You are still in one piece, aren't you?

"I am not crying for myself, my lord, I weep for you," the clown explained.

"For me?" exclaimed the Caliph in surprise.

"Certainly! If I get beaten so badly for having been only a few minutes on the throne, how many beatings must you have suffered in all the years that you have been there?"

Original source: 366 and More Fairy Tales, 1990.

In the beginning it is stated that the aims of the lesson are to read a narrative text and to identify the characteristics of a narrative text. The questions provided for this text are:

1. Identify the introduction of the story (UM – identify text development)
2. Mention who? When? Where? (UM – identify character(s) and answering specific text-based questions)
3. Mention the action verbs and words that show time (UM – identify (parts of) text development)
4. Mention its conflict (RM – drawing conclusion/inferring sequences)
5. Write how the conflict is resolved (RM - drawing conclusion/making inferences).

Similar to the previous text, the questions in this text also vary from easy to more complex ones as shown in no 4 and 5. The questions can also be expanded on finding main ideas or lesson learnt from the story and also evaluating characters.

REFERENCES

Cahyono, B.Y. (2008). The continuous improvement learning programme for English for English language teachers: An Indonesian experience. In M.E. Vethamani & M.K. Kabilan (Eds.), Practices and issues in English language teacher development (61-78). Petaling Jaya, Malaysia: Sasbadi Sdn. Bhd.

Hanafie, I. (2007). Plus minus kurikulum tingkat satuan pendidikan (KTSP). http://meysiska.blogspot.com/2009/05/plus-minus-kurikulum-tingkat-satuan.html. accessed online Nov 13, 2009.

Jazadi, I. (2003). An investigation of current constraints and potential resources for developing learner-centred curriculum frameworks for English language at high schools in Lombok, Indonesia. Unpublished PhD thesis, Division of Education, Arts and Social Sciences, University of South Australia, Adelaide, Australia.

Kistono, Andayani, E.T., Ismukoco, Tupan, A. (2007). The bridge English competence 1 for SMP grade VII. Bogor: Yudhistira.

Kistono, Andayani, E.T., Ismukoco, Tupan, A. (2007). The bridge English competence 2 for SMP grade VIII. Bogor: Yudhistira.

Lampung Post (6 Desember 2006). Kurikulum: pemahaman KTSP bervariasi. http://www.lampungpost.com/cetak/berita.php?id=2006120601353359 accessed online Nov 13, 2009.

Prasetiyo, B. (2009). Mengefektifkan kemampuan guru dalam menerapkan KTSP melalui MGMP Swadaya. Jurnal Penddikan Inovatif, volume 4 number 2, pp. 48-53.

Priyanto, A.D. (2009). Maximizing SFL Contriburion to ELT Indonesia. UAD TEFL Conference. Yogyakarta: UAD Press.

Rosenshine, B.V. (1980). Skill hierarchies in reading comprehension. In Spiro, R. J., Bruce, B. C., Brewer, W.F (eds). Theoritical Issues in Reading Comprehension; Perspectives from Cognitive Psychology, Lingustics, Artificial Intelligence, and Education. New Jersey: Lawrence Erlbaum Associates, Publishers.

Suara Merdeka (16 Sept 2009). Belum semua sekolah terapkan KTSP. http://suaramerdeka.com/v1/index.php/read/cetak/2009/09/16/80739/Belum.Semua.Sekolah.Terapkan.KTSP. acsessed online Nov 13, 2009.

Sulistiani, Endang. 2009. The Perception of English Teachers at Public Junior High Schools in Malang on the Implementation of School-Based Curriculum (SBC). Thesis, English Language Education, Faculty of Letters, State University of Malang

Wardiman, A., Jahur, M.B., Djusma, M.S. (2008). English in focus 1: for grade VII Junior High School (SMP/MTs). Jakarta: Pusat Perbukuan, Departemen Pendidikan Nasional.

Wardiman, A., Jahur, M.B., Djusma, M.S. (2008). English in focus 2: for grade VIII Junior High School (SMP/MTs). Jakarta: Pusat Perbukuan, Departemen Pendidikan Nasional.

BECOMING AN EFFECTIVE TEACHER OF READING COMPREHENSION

[Teachers play very important roles in any educational improvement. Therefore, it is crucial to understand pedagogical actions that teachers should do to help students take charge of the learning process. With respect to the teaching of reading comprehension in JSS, this document attempts to offer research-based practical and observable strategies that teachers can do in order to improve their teaching quality. It is to accompany a document on the elaborated standards, both of which are to be used in Teacher Professional Development Program in my PhD research project]

Siti Nurul Azkiyah – UIN Syarif Hidayatullah Jakarta
GION, University of Groningen, NL
[July 2010]

Special thanks to Prof. M.P.C. van der Werf, Dr. Simone Doolaard, Prof. Bert Creemers for supervising the development of this document.

Table of Content

List of Table

List of Figure

Introduction

This document will discuss strategies of teaching reading comprehension based on previous research found to lead to better students' outcomes. The introduction will firstly describe some teacher-student activities in the classroom followed by a discussion on effective teacher characteristics. The second part will, moreover, explain the application of effective teacher characteristics in the teaching of reading comprehension. Finally, the document provides an example of effective reading instruction.

The followings are some activities we may find in English classrooms. Think about the purpose of the activity.

Teacher 1: Assalamu'alaikum Wr. Wb. Good morning students! Last time we have learned about famous people. Anybody remember what we have learned in our previous lesson?

Students: [silent]

Teacher: Rita, do you remember one of the famous people?

Rita: Thomas Alpha Edison.

Teacher: Very good! Do you still remember his invention, Rinto?

Rinto: Ehm….[silent for quite long time]

Teacher: [the teacher did not say anything but pointed to the light bulb to give a clue to Rinto]

Rinto: He find electricity.

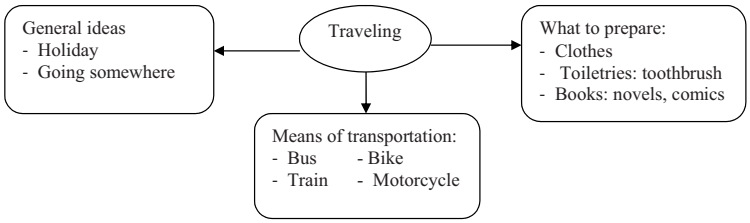
Teacher: Excellent Rinto. He found electricity. We use past tense since it happens in the past.

Teacher 2: Today we are going to learn about travelling. When you have holidays for instance, do you visit your grandmother or grandfather in other cities?

Students: Yes!

Teacher: So, traveling is a common activity in our life. We will read a descriptive text about traveling. When we read, it is common that sometimes we do not understand all words mentioned in the text. Therefore, we will learn how to recognize words or paraphrase meaning in the text, and then we will learn how to find the main idea of the text and its paragraphs.

Now let's share what you have in mind when you hear the word "traveling". This is one of the ways to help us guess what words will be there in the text and eventually what the text will be about. I'll draw a picture to map what we already know about traveling. [the teacher writes down students' answers]



Teacher: anybody wants to start? What is in your mind when you hear the word “traveling”.

Agung: Holiday, Mam!

Teacher: Ok, holiday. We have mentioned about it before. Others?

Dina: We go somewhere…

Teacher: Good, we go to certain places. What do we prepare when we want to travel, for instance when you have holidays and you want to go and stay in your grandma’s house for few days?

Students: [silent]

Teacher: Do you prepare your clothes, Dewi?

Dewi: Yes, Mam.

Teacher: Good, Dewi. Do you prepare any other thing, students?

Udin: Ehm…..toothbrush and may be books.

Teacher: Excellent. You may want to read your books. You can bring novels or comics or even your textbooks. What about the means of transportation? How do you travel?

Ridho: by bus.

Teacher: Ok, Ridho, good!. Do you go only by bus? Girls, how do you travel to your grandma?

Rahmah: by train.

Teacher: Thank you Rahmah. No body goes by bike or motorcycle with your parents?

Ana: I do, Mam. My grandfather is very close to my house. I go by bike.

Teacher: Very good, Ana. So we can travel by various means of transportation. Now, let’s read the text.

Teacher 3: Anybody knows how to find the main idea of a paragraph? Today we will learn this skill again. Since, we have learned this skill in the previous session I like to invite you to explain the strategies of finding the main ideas of paragraphs. Joko, can you please start the first tip that we can use to find the main idea of a paragraph?

Joko: Look at the first or the last sentence of the paragraph.

Teacher: Very good, Joko. The main idea of a paragraph can take place anywhere in a paragraph, but it is quite common that it is situated in the first or the last sentence of a paragraph. What if the first or the last sentence is not the main idea? Any volunteer?

Tina: It can mean that the main idea is implicit. We have to read all sentences and summarize.

Teacher: Excellent Tina. Now let's do the exercise. I have some paragraphs. You can work in pairs and find the main idea of each paragraph.

Teacher 4: Today we will learn how to make inferences. I'll give you some tips, which I hope can be used to answer the exercises that you will do.

Teacher 5: Ari, can you please answer question no 5? [silent]

[here is the text]. Kelik read a book he liked a lot. He was really sorry when the story ended. He thought the author had done a great job. The next day Kelik took the book back to the library.

[here is the question]. What author might Kelik look for when he borrows another book from the library?

Ari: [silent for quite long time]

Teacher: Ok, Ari, I think the wording might be quite complicated for you, so I will state the question in a different way. I am sure you can answer it. What kind of book Kelik will borrow after he returns the book? Or….which author Kelik will look for when he comes to the library?

Ari: [still silent]

Teacher: Ari…for instance, today you eat banana and you think you like it. Will you eat banana again tomorrow or next time?

Ari: Yes!

Teacher: Now….Kelik liked the book a lot. He wants to borrow a book again in the library. Will he borrow similar book? Or will he borrow books written by the same author?

Ari: Yes

Teacher: excellent Ari. Others….do you agree that Kelik will look for the same author when he comes to the library?

Teacher 6: Andi, why do you think the main idea of the second paragraph is the first sentence?

[Here is the paragraph] There are three kinds of travel books. The first are those that give a personal account of travels. If they are informative and have good index, they can be useful to you when you are planning your travels. The second are those which purpose is to give an objective description of things to be done and seen. It can be classified as selective guidebook. The third kind is those books which are called 'a guide' to some places or others. If they are good, they will give an analysis or an interpretation. For instance is the first ones. They are inspiring and entertaining. But their primary function is to assist the reader who wishes to plan in the most practical way.

Andi: Usually the main idea is in the first paragraph.

Teacher: Andi, do you think the other sentences support the first sentence?

Andi: Yes, Sir.

Teacher: Good, Andi. Lala, can you mention the supporting sentences?

Lala: The second sentence, "The first are those that give a personal account of travels"

Teacher: Very good, Lala. Others…..can you mention other supporting sentences?

Iwan: The second is those which purpose is to give an objective description of things to be done and seen. It can be classified as selective guidebook.

Teacher: Bravo, Iwan. And the last one is…..

Students: [read the sentence together] The third kind is those books which are called 'a guide' to some places or others.

1. Effective Teacher Characteristics

Why do teachers do such activities as described in the introduction? Do they represent effective teachers? They, indeed, apply some factors that have been found to lead to better students' outcomes. Teachers have been proved to have the biggest influence on students' outcomes and therefore there has been research aimed at understanding teacher pedagogical actions at the classroom level which lead to better students' outcomes. These pedagogical actions refer to observable teacher instructional roles, which can be practical for teachers especially when they like to improve their teaching quality.

Research has found there are at least 8 aspects that teachers can use as a reference concerning actions teacher can do in their classroom to improve their teaching practice (Creemers &

Kyriakides, 2008). They are 1) orientation; 2) structuring; 3) teaching modeling; 4) application task; 5) questioning; 6) building classroom as a learning environment; 7) management of time; and 8) assessment. What are they and how do they work in real classroom teaching and learning? The following explanation on the examples described in the introduction might answer the question.

Teacher 1 opens and starts the lesson by ***reminding and asking students*** about previous lesson. This activity can be considered as a review of previous lesson, which is important not only to remind students about previous lessons but also to connect the day lesson with what students might have known from previous lesson. Thus it helps activating students' previous lesson, which has been considered as one of the characteristics of effective teaching.

Teacher 2 attempts to do at least three activities. The first is ***structuring*** in which he/she mentions explicitly about the topic, which is travelling. He/she then moves to ***orientation activity*** by connecting the topic with students' own experience. This, again, is helpful in activating students' previous knowledge and also in explaining the importance of studying certain topics or skills. Furthermore, the teacher continues with structuring activity, in this case, explaining specific reading skills students will learn. This is useful to help students aware about the competences and activities they will do / learn during the day lesson. Moreover, the teacher also explains the reasons why students will learn those skills and it is expected that by doing so, the activities and the skills will be more meaningful to students. It is also suggested that teachers encourage their students to find out themselves the importance of studying what they are going to learn. Finally the teacher provides ***modeling*** (the use of strategies to solve problems) activity to recognize words which may appear in the text. This is similar to prediction, in this case predicting words and subtopics (such as general idea of traveling, what to prepare and means of transportation) in the text, which eventually will lead to understanding the whole content of the text.

Teacher 3 asks his/her students if they know the strategies or the way to find the main idea of a paragraph. This activity is also ***modeling*****.** It is expected that effective teachers help their students to use strategies and/or develop their own strategies for them to solve different types of problems (in this case exercises). This can be useful to promote self-regulated learning. In the example, teacher 3 attempts to ask students to present the strategies as the skill has been taught and he/she expects that the students know already the strategies. Thus, it is possible that the students provide the learning strategies. It is also possible that the teacher provides the learning strategies as shown in the activity carried out by **teacher 2 and 4**. It is also important to note that this teacher also mention ***application*** task students are supposed to do. Application task is exercises that teachers provide for students to practice strategies they learn.

Teacher 5 asks a student to answer a question. Knowing that the student keeps silent for quite a long time, the teacher tries to rephrase the question into easier words hoping that the student will be able to give the correct answer. He/she, furthermore, provides clue to enable his/her student to answer the question. This activity is called ***questioning*** and effective teachers are supposed to provide relevant questions as well as constructive feedback or reaction. When the questions are

difficult for instance, teachers should give more time, rephrase them into easier words or provide clues. Teacher 5 is very good at giving feedback and reaction, which is very useful in **developing classroom as a learning environment**. In addition, it is also suggested that teachers provide both process and product questions. Process questions require students to go beyond the printed information in the text. They may be asked to explain why they come to certain answers. The question raised by **teacher 6** is an example of how process question is raised. Product questions are those that require students to recall facts, concepts, or procedures mentioned explicitly in the text.

What about the other three factors? Do teachers also apply them? Since the examples above are only fragments of certain activities happening in the classroom, they do not really give us information on how teachers manage all activities in one lesson, both in terms of engaging all students and also in managing time. However, most teachers in the example attempt to engage different students to be on task. Teachers assign different students to answer different questions. Teachers also seem to remember the names of their students well, which is very good. In addition, teachers also provide relevant feedback to students' answers or reaction in order to praise or to provide constructive reaction (for instance to give a clue or to change a question into easier words). All of these activities are useful in **developing classroom as a learning environment**. It is found out that teachers who provide positive and constructive feedback have better students' outcomes.

Unfortunately, we do not know teachers' management of time. The most important aspect in managing time is that teachers maximize students' learning time during the lesson. Lastly, assessment should be seen as an integral part of teaching. Information gathered through student assessment should enable teachers to identify their students' needs as well as to evaluate their own teaching practice. It should be noted that assessment can take place at any time during the lesson including at the end of the session. In the examples, for instance teacher 1 and 3, assessment of previous lesson takes place in the beginning of the lesson. Furthermore, teacher 2 and 5 gives clues to students, which the same time also assess students and then provide necessary reaction to the result of the assessment. Since they know that students still have difficulties in answering the questions, they attempts to provide clues to enable students find the correct answers. Thus, they use the information gathered during assessment to provide their students' needs.

These eight factors do not refer only to one approach of teaching such as the direct teaching model or the constructivist approach. The direct teaching model for instance appears on the structuring and questioning, whereas orientation and teaching modeling represent constructivist approach of teaching. In addition, collaboration and grouping technique are also promoted through application, and teacher – student and student – student interactions are indicated in the role of teacher to create classroom as a learning environment. Furthermore, opportunity to learn and time on task are considered as the most significant effectiveness factors and therefore teachers are expected to organize and manage the classroom in order to maximize engagement rates.

These eight factors can serve as guidance in how the teaching should be delivered. Further concrete steps concerning what teachers are supposed to do with respect to the teaching of reading comprehension are provided in table 1.

Table 1 Effective Teacher Characteristics and their concrete activities

FACTOR	ACTIVITIES
ORIENTATION	1. Present or invite students to present the aims of studying certain topic/theme, type of text and specific reading skills (as stated in the elaborated standards) 2. When presenting the aims, some aspects should be considered: a. The aims could be linked to previous activities, students' daily life, social function of the theme and/or the text b. The presentation of the aims could take place in different stage c. The presentation of the aims should be clear for students d. If possible, the presentation of the aims should consider different ability or learning needs of students
STRUCTURING	1. Present the structure of the lesson, which, in the case of reading comprehension should include: - Topic - Type of text - Specific reading skills to be taught 2. Explain the link among different activities that students will do during the day lesson 3. Signal the transition between one phase/activity and the others and review or mention the aims of the next activity 4. When presenting the structure, please bear in mind that it should be clear for students. In addition, if possible, the presentation of the structure should consider different academic background of students. Thus, the presentation could be verbal or written or presented through a power point presentation.
MODELING	1. Present the strategies of specific reading skills planned to be taught. 2. When possible, it is suggested that teachers invite students to present the strategies of the skills planned to be taught. 3. When it is not possible to ask students to present the strategies, teachers can at least engage students in developing or demonstrating the strategies. 4. When presenting the modeling activity, clarity of the activity for the students should be considered. If possible, different types of modeling could be provided to address different ability or learning needs of students.
APPLICATION	1. Prepare tasks or exercises of each specific reading skills planned to be taught to apply the strategies presented during modeling. 2. When developing the tasks, it is suggested to consider the followings: a. Whether the application tasks should replicate the same activities presented during modeling.

APPLICATION	b. Whether the text used in modeling should be different from the text used in application activities. c. Whether the same text should be distributed to the whole students in the class.
QUESTIONING	1. Prepare questions and guide classroom discussion through necessary questions in accordance with the skills planned to be taught. It is suggested to begin with questions which help students recall previous lesson. 2. Both process and product questions should be balanced. Process questions are those attempting to require students to explain how they find their answers, whereas product questions are limited to ask students to get the answers from the text they read. 3. Think about appropriate reaction when no answer is given by students. 4. Provide positive and constructive comments or feedback to students.
ASSESSMENT	1. Think and prepare strategies and tool to make sure that students understand the main ideas or skills of the day lesson. 2. It is possible that the assessment is carried out when teachers pose questions to the students. 3. Think about the time to deliver assessment. 4. The result of the assessment should enable teachers to improve their teaching practice.

2. The Application of the Effective Teacher Characteristics in the Teaching of Reading Comprehension

The effectiveness enhancing factors presented in table 1 have been argued as generic teaching skills. However, these factors will require teachers to understand the content of the subject. Therefore, both the effective teacher characteristics and the strategies of teaching reading comprehension are discussed in this document. The eight factors serve as a framework to guide how to deliver teaching learning activities in the classroom in order to achieve quality teaching and finally the educational standards. The discussion on teaching reading comprehension especially concerning the skills elaborated in the standards are expected to provide a reference for teachers when they have to present modeling or provide application activities as well as questioning and assessment.

This section will, furthermore, provide specific teaching strategies for the reading skills elaborated in the standards of content (the summary of elaborated standards is provided in the appendix of this document). It could be useful especially when teachers want to provide modeling or develop application tasks as well as questions to reach the intended competence or skills.

Understanding meaning

1. Understanding words in context.

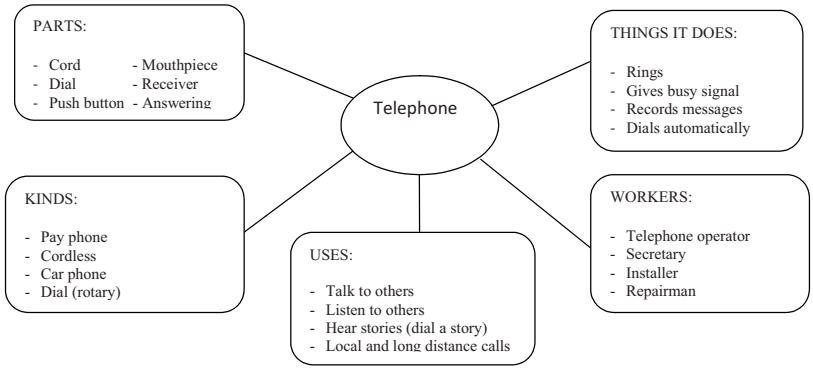
The specific competence under this category is r**ecognizing words / paraphrase meaning**. This can be considered as pre-reading activities as it is prerequisite for readers to have this ability in

order to achieve comprehension. Therefore, by itself, it is an insufficient goal of reading instruction. There are some ways to teach this skills, two of which will be offered in this document.

Firstly, teachers are expected to find key words in the text and relate them to students' lives. It will be easier and useful when teachers can visualize or bring real objects representing both key words and students' lives. Pictures or real objects are worth a thousand words and visual display helps readers understand, organize, and remember some of those thousand words. This is important not only to call for students' background knowledge but also to develop students' interest and finally to raise students' awareness that what they are reading are related to their lives and therefore meaningful to them.

Secondly, this sub-skill is closely related to the development of vocabulary, which is an important contributing factor to reading comprehension. There are different strategies of building vocabularies. Semantic mapping has been found to be an effective strategy. It is easy to prepare and does not need any aids of technology, and thus is very useful for Indonesian classrooms in general. Teacher can identify sub-topics of the texts and together with students they can brainstorm words students know about the sub-topics. The following is an example of semantic map for "telephone".

Figure 1 Semantic Map for Telephone (Pittelman & Heimlich, 1991, p. 46)



2. Locating details. Specific reading skills under this category include a) reading pictures, tables and number; b) answering specific text-based questions; c) identifying and describing characters; and d) identifying types and pattern of text structure/development.

2a. Reading pictures/tables/numbers. Pictures, tables, and also numbers are common in our daily life. Besides, children also like pictures and therefore it is important to teach our students on how to understand them. There are at least three points for this skill. The first is to identify the meaning of pictures and signs such as those to indicate that parking and smoking are not allowed, to indicate occupation and then moves gradually into more complex picture or sign as well as exercise such as developing narrative story based on provided pictures. The second is the ability to locate information from a text into a table or vice versa and the last one is to identify the meaning of words related to number such as half, some, more and so forth.

It will be helpful when teachers could bring big pictures or signs or even real objects and then brainstorm with the whole class of possible meanings the pictures or signs may convey. Students may have some background knowledge and thus it will be helpful to begin by asking students if they have come across similar pictures or signs to recall what they may have had in their mind. Concerning the skill to locate descriptive information into a table, teachers are expected to begin and model easier exercise such as locating the schedule of students' daily activities into a table and vice versa and then move to more complex activities. Finally, in order to help students to identify the meaning of words related to number, teachers can bring a round block which can be divided into small pieces to show students how much is half, some and so forth. When the class or the teachers are facilitated by computers or laptops, the teacher can make use of slides which show the division. Teachers can also use available sources such as the white or black board or even paper and draw the division. Making this visual for students will be more useful because students remember pictures more than words.

2b. Answering specific text-based questions. It is a literal comprehension which questions usually refer to W-H (what, when, who, where, why and how) questions. This type of question is dominant in both textbooks and national exams. Teacher should previously identify important information from the text and create questions that require students to identify the information. The questions may overlap with reading pictures/numbers/table, identifying characters, identifying types and pattern of text organization.

To make the classroom more interactive, teachers can also group students and distribute cards (which have to be prepared in advance) of what, when, who, where, why and how and ask each group to find the answers for each card. When they finish, teachers can facilitate discussing the answers together with students.

2c. Identifying and describing characters. In a narrative text, the presence of characters is usually dominant. Several authors state that in reference to characters in the text, most teachers rely on two questions. The first is "who is the main character(s) in the story" And the second is which character you like or dislike most." These two questions are also dominant in Indonesian books although they also require students to describe the characters. It is suggested that teachers can gradually lead students to understand who the characters are and what they do to how and why characters think, feel, and act as they do. For example teachers can ask a student to mention

the main character of a story and then ask other students to mention three or five things about the main character and finally ask them the reason why the character is such. Teachers can do this to the whole class or in groups. If teachers decide to do this activity in groups, they can make use the strategies of grouping in Peer Assisted Learning Strategies (PALS) in which students are grouped in pairs and each partner will read one paragraph and retell what each has red.

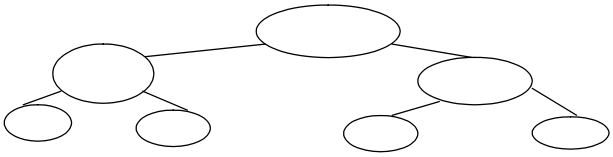
2d. Identifying types and pattern of text structure/development. Text structure and students awareness of text structure are highly related to reading comprehension. It is important to explain to students in the beginning that writers use text structures to organize information. Therefore, instruction should help students to identify the physical textual organization of texts such as headings and subheadings and explanation of their purposes, signal words, topic sentences and also a description of where they usually occur in a text. Another important point is to identify patterns of text organization, which can include cause and effect, problem and solution, comparison and contrast, description, classification, analysis, argument and evidence, procedural sequence, chronological ordering and the like. The introduction of this text structure is important since students can also apply their knowledge and skill when they read different texts. In addition, the skill can also be used when students have their writing classes since they should enable them to organize and edit information necessary for composition.

The Indonesian standards of content introduce five different types of texts for JSS: descriptive, procedural, recount, narrative and report. For grade VIII as the focus of the study, the texts are limited to **descriptive**, **recount**, and **narrative**. Descriptive text, in a simple way, can be described as a text that describes the feature of someone, something or certain places. It is characterized by the use of the simple present tense, relational clauses, adjectives, and so on (Agustien, 2006). Whereas recount is a text that reports events in order to inform or entertain readers. It begins with orientation to explain the factual information (what, who, when, and where) and then the sequence of events and finally personal comments on the event. Narrative text usually is comprised of elements which can be referred to as story structure or story grammar. Furthermore, when readers read narrative text they expect character development, conflicts, episodes, and conclusions. What makes narrative different from recount is the presence of complication, problems, conflict, moral values. Recount focuses merely on the series of events themselves.

When teachers want to focus on this skill, they can use reading guides, story maps, or graphic organizers (Venn diagrams, matrices, flow charts) indicating the structure of different text and ask students to complete them. Research has shown that in L2 context graphic organizers have proven to be effective not only in helping students recognize text structure but also in highlighting main concepts and their relations with supporting information. The followings are examples of text structures that indicate description / classification (for descriptive text), time line or sequences of events (for recount) and narrative.

Figure 2 Text Structure to Indicate Description / Classification, Time line / Sequences of Events and narrative structure (Grabe, 2009, p. 262 – 264)

Description / classification



Time line / sequences of events

T1 ----------
T2 ----------
T3 ----------
T4 ----------

Narrative structure

- Description of setting
- Characters development
- Conflict
- Episode sequences: conflicts and conclusion

Responding to Meaning

3. Finding topic, main ideas / predicting titles. Titles can usually tell about the general topic of texts. Some writers call the term as the global main idea. Whereas those situated at paragraph levels are the local main idea, which could be explicit and implicit. A text usually has several local main ideas but only one global main idea that governs the entire text and receives support from all local main ideas in the text. However, the global main idea is not the sum of all local main ideas put together. Instead, it is formed through a hierarchically structured text. Therefore, knowledge of text structure would help readers to identify the global main idea. It is a highly recommended reading strategy in finding global main idea. Thus, the teaching of this sub-skill can be combined with the teaching of recognizing text structure.

Furthermore, with reference to the local main ideas, the implicit ones can be identified from their location. Instruction should begin by using a paragraph that has an explicit main idea, which usually appears in one complete sentence, is related to the majority of sentences in the paragraph and may occur in any point in a paragraph. Students learn to identify main idea statements that are embedded in paragraphs by asking themselves if each sentence is relevant to the first or the last sentence. If not, it could mean that the main idea is implicit and students have to develop the

main idea statement. In this case teachers have to provide exercises to find both explicit and implicit main idea statements (Dickson, Simmons, Kameenui, 1995).

Study guide could be useful to lead especially poor readers to identify what are important to read to find the main idea. Poor readers usually get lost in the details and therefore study guide will help them focusing on what to read. He gives an example of study guide.

Table 2 Sample from a Study Guide for Underachieving Readers (Askov, 1991, p. 14).

Read the last paragraph. This summarizes the text.
1. Read the first paragraph and write the main idea.
2. Skip paragraph two.
3. Read the first sentence of paragraph three. How does this help you understand the main idea in the first paragraph?

4. Drawing conclusion. There are four sub-skills categorized as drawing conclusion in the elaborated standards. They include a) finding author's purposes, attitude, tone, or mood; b) drawing inferences, c) inferring sequences; and d) evaluating ideas in the text.

4a. Finding author's purposes, attitude, tone, or mood. Writers, for sure, have intention to write, which will be matched with the way they write. It is the readers' job to find the purpose of the writer. When, for instance, they want to explain the procedure of working with computer, they may use procedural approach. However, when they want to tell stories they may use a narrative structure. In addition to looking at the general ideas presented in the text, the purpose can also be seen from the way the texts are structured. Thus teachers can lead students to look at the general ideas, the structure of the text which includes signal words and stages of the text development as previously explained in *identifying types and pattern of text development*. Furthermore, attitude, tone or mood is usually associated with positive or negative reaction of the writers towards the issues that they bring in the text. In identifying author's attitude, tone, and mood, teachers can explain to the students that they need to find key words representing feeling or reaction. *Questioning the Author* approach, which is dominated by teacher-student and student-student discussions about the text, could be useful to teach students about the skill. Teachers use queries to broaden and deepen students' thinking about the writers' responses towards the issues they explain in the text.

4b. Making inferences. In a simple way, to infer means to understand or to interpret unstated but implied information using available information in a text. Inference can be as simple as associating the pronoun of "she" with previously mentioned female person. However, in the elaborated standards, this simple pronounce reference is categorized as recognizing words / paraphrase meaning. Furthermore, inference can of course be complex that requires readers to draw conclusion aided by their own background knowledge as well as explicit information in the text.

There are some steps to help students identify unstated but implied information in the text they read. Beforehand, it is important to explain to students that many times writers leave some ideas unstated in the text and it is the readers' job to find the ideas. Those unstated ideas can include *people, animal, things, places, time, actions,* and *feeling.* The followings are an example and steps to figure out unstated ideas.

Table 3 Example of an Inference Question

Dewi was walking home from school. She passed a small river. When she was home, her mother asked her to put her wet shoes on top of the rack.

How did Dewi get her shoes wet?

Whereas the steps are:

Table 4 How to Figure Out Unstated Ideas (Baumann, 1991, p. 68)

1. Read and understand the facts
2. Decide what the writer has left out
3. Figure out what the unstated ideas are
    a. Think about what you know about the story
    b. Look for clues in the story/text
    c. Make a guess about what the unstated ideas are
4. Read on to check and see whether you were correct
5. Go back to step 3 if you were wrong

4c. Inferring sequences

Inferring sequences can include inferring sequences of events, inferring comparison, as well as causal relationship. They are briefly explained in the following.

Inferring sequences of events: in this case, students may be requested to conjecture as to what action or incident might have taken place between two explicitly stated actions or incidents, or they may be asked to hypothesize about what would happen next if the text or the story had not ended as it did but had been extended.

Inferring comparison: the students are required to infer likeness and differences in characters, times, or places. Such inferential comparisons revolve around ideas such as: "here and there," "then and now," "he and he," "he and she," and "she and she."

Inferring cause and effect relationships: the students are required to hypothesize about the motivations of characters and their interactions with time and place. They may also be required to conjecture as to what caused the author to include certain ideas, words, characterizations and actions in his writing.

4d. Evaluating ideas presented in the text. Evaluation is a high order thinking which can be difficult especially when the information is not explicit. Looking at the Indonesian classroom and national exam, evaluating ideas vary from judging that certain statements are true or false according to the text or whether certain statements exist in the text. In order to be able to evaluate ideas, students need to understand both general and detailed ideas presented in the text. Therefore, this skill should be taught together with other skills especially identifying main ideas and text structure / development.

5. Predicting outcomes. In order for the students to be able to predict outcomes, teachers can ask students if they have ever been able to figure out what a text is about by reading its title. Teachers can then distribute different titles of text and ask students to predict the content of the text. After the brainstorming, teacher can distribute the texts and asks students to check if their prediction is correct. Another aspect of predicting outcomes is to predict what the next paragraph will be after reading or discussing previous paragraph. This will be useful when reading stories or narrative text. Still another aspect of this category is to predict responses of given situations. This skill is examined in the national exam throughout the years and thus it is suggested that teachers spend some time for this skill. Students can be grouped and assigned to give responses to various situations.

3. Designing Reading Instruction

We have discussed the effective teacher characteristics and the application of those characteristics in the teaching of reading comprehension. We have also discussed specific teaching strategies of reading skills mentioned in the elaborated standards which could be useful especially when teachers want to deliver modeling or develop questions and application tasks. The next question is what the instructional design looks like?

The following is an example of a reading instructional design, which is made in line with the procedure of writing lesson plan required in the Indonesian curriculum.

Name of school	: School A
Subject	: English
Class/semester	: VIII/1
Theme	: Famous people (Bung Hatta)
Skill	: Reading
Time allocation	: 2 x 40 minutes

Standard of competence:

- Understanding meaning in short and simple descriptive text to be able to interact

Basic competencies:

- Meaningful reading aloud short and simple functional text and essays
- Responding meaning to short and simple functional text
- Responding meaning and rhetorical stage in short and simple essays

Specific Reading skills:

Understanding meaning

- Recognizing words / paraphrase meaning
- Identifying and describing characters
- Answering specific text-based questions

Responding to meaning

- Finding topic / main ideas

Assessment checklist :

- Students can identify the meaning of most (80%) of words / phrases in the text.
- Identify pronounce reference.
- Students can identify all persons/characters mentioned in the text.
- Students can describe the feature or characteristics of the persons mentioned in the text.
- Students can identify factual information explicitly stated in the text. Usually the questions include 5W I H (who, what, when, where, why, and how).
- Students can identify the most important message or main idea of the text.

The text: English on Sky VIII – published by Erlangga p. 176

Muhammad Hatta was one of Indonesian founding fathers. He lived from 1902 until 1980. Together with Soekarno, he proclaimed the independence of Indonesia on 17th August 1945.

Muhammad Hatta was born on 12th August 1902 in Bukittinggi, West Sumatra. When he was still in Junior High School in Bukittingi he joined the league of Young Sumatrans.

When he finished his study in Bukittinggi, he moved to Batavia. Then, he went to the Netherlands to continue his study. When he was there, he participated actively in the National Movement. As a result, he was arrested by the Dutch government.

In 1932, Bung Hatta went back to Indonesia. He joined a political organization called *Pendidikan Nasional Indonesia*. This organization wanted Indonesian people to know many things about politic. Because of this activity, he was arrested again. He was sent to Boven Digul and later to Banda Neira as a prisoner. Before the Japanese invaded Indonesia in 1942 he was brought to Java.

On 17th August 1945, two days after Japan surrendered to the Allies, Bung Karno and Bung Hatta declared the independence of Indonesia. Then, they were selected as the president and vice president. Bung Hatta was the vice president until 1956. He resigned and concentrated in writing. On 14th March 1980 Bung Hatta passed away in Jakarta.

Muhammad Hatta was one of the greatest people in Indonesia. People will always remember him as an honest and sincere person. (237 words)

Preparation activity

1. Read the text and identify key words in the text.
2. Think about the importance of reading this text for students; relate the topics with students' life or previous lessons.
3. Create relevant questions and tasks according to the specific reading skills planned to be taught.
4. Plan the activities and how the activities are going to be organized (whether to use whole class instruction or grouping students).
5. Decide how much time needed for each activity.

Table 5 Classroom activity

Activity	Approach / technique
Opening, introduction (5 – 10 minutes)	Greeting, calling for students' attendance, checking and discussing HW.
Structuring (5 minutes)	Mention explicitly to the students about the day lesson: - Topic: famous people in Indonesia - Type of text: descriptive - Specific reading skills students will learn a. Recognizing words / paraphrase meaning b. Identifying and describing characters c. Answering specific text-based questions/identifying factual information from the text d. Finding topic / main ideas When presenting the structure, please bear in mind whether to mention them orally or write them down on the board or in other ways, which are better understood by students. It is suggested that teachers explain that recognizing words/paraphrase meaning is a prerequisite to reading comprehension, and therefore they are going to learn it before proceeding to the next reading skills. Another explanation is the fact that the text is a descriptive text and in order to understand the whole text, identifying factual information including the characters mentioned in the text is useful. So, after recognizing words, students will learn identifying and describing characters, identifying factual information and at the end finding topic or main ideas.
Orientation (3-5 minutes)	Before proceeding to the main activities, it is important to present or invite students to explain the importance of reading text about famous people in Indonesia. Teacher can initiate the explanation by saying: "famous people usually have strong characters that make them different from others. Any body wants to be famous? If we want to, we can learn how to build our characters by reading their biography. Now we have a short text about Bung Hatta. I am sure all of us are familiar with his name. So, let's read the text and see what we can learn from him".

Orientation (3-5 minutes)	Another way is to brainstorm on what students know about Muhammad Hatta (Bung Hatta). Teacher can bring the photo of Bung Hatta and brainstorm what students know about the picture. Teacher can use semantic map to write students' answers. This is similar to pre-reading activities in reading theories, which is useful to activate students' background knowledge. After brainstorming, the teacher can ask students about the importance of reading such text.
Structuring (1 minute – signaling transition	Teacher can then signal the transition that the reading activity will begin with reading aloud together and that he/she will model the reading before asking students to repeat after him/her.
Modeling 1 (reading aloud – 10 minutes)	Teacher can first read the text or ask a student to read sentence by sentence moving from one student to another student. In this stage, teacher should pay attention to the pronunciation. Mispronunciation should be used as a starting point to ask all students to repeat the correct one. This should not last for more than 10 minutes
Structuring (1 minute)	Teacher can explain that the next activity is recognizing words/paraphrase meaning.
Modeling and application (30 minutes)	Using the semantic map drawn during orientation stage, teacher can further discuss the key words and make sure that most students understand most words in the text. The activity should then move to the main reading skills planned in the first structuring stage and teacher is expected to signal the transition and review the aims when necessary. The main activities will include: a. Identifying and describing characters b. Answering specific text-based questions/identifying factual information from the text c. Finding topic / main ideas Teachers will have to prepare questions/tasks to cover all the above skills. Students can be grouped and each group will have different sets of questions. It should be noted that the questions for point a and b could be overlapped.
Assessment (7 minutes)	Possible assessment activity (could also considered as post reading activity) for this text can include: a. Asking students to write what they think or feel about Bung Hatta (in one or two sentences for 2 -3 minutes) and collect their sentences. b. Asking few students to read their sentences.
Closing (3 minutes)	Teacher can summarize what students have learned during the day lesson and explain the next lesson. Teacher should clarify to students if they have any questions regarding the day lesson.

References

Agustien, H.I.R. (2006). Competence-Based Curriculum and its practical implications. Paper presented at UTM International Seminar, Bandung, 22 February 2006.

Askov, E.N. (1991). Teaching study skills. In Hayes, B.L. (eds). *Effective Strategies for Teaching Reading*. Boston: Allyn & Bacon

Baumann, J.F. (1991). Teaching comprehension strategies. In Hayes, B.L. (eds). *Effective Strategies for Teaching Reading*. Boston: Allyn & Bacon

Creemers, B.P.M. and Kyriakides, L. (2008). *The Dynamics of Educational Effectiveness*. London: Routledge.

Dickson, S.V., Simmons, D.C., Kameenui, E.J. (1995). *Text organisation: curricular and instructional implications for diverse learners*. EDRS.

Grabe, W. (2009). Reading in a second language; moving from theory to practice. Cambridge University Press.

Pittelman, S.D. & Heimlich, J.E. (1991). Teaching vocabulary. In Hayes, B.L. (eds). *Effective Strategies for Teaching Reading*. Boston: Allyn & Bacon

APPENDICES FOR CHAPTER 4

Appendix 4.1

The Differences in Teaching Quality among the Three Groups from Measurement to Measurement (*t*-test)

Group	M1 to M2	M2 to M3	M1 to M3
Observer data			
Experimental one	M1 = 1.80 (.26) M2 = 1.74 (.36) *t* (15) = .79, *p* = .44	M2 = 1.74 (.36) M3 = 1.76 (.22) *t* (15) = -.23, *p* = .82	M1 = 1.80 (.26) M3 = 1.76 (.22) *t* (15) = .57, *p* = .58
Experimental two	M1 = 1.94 (.37) M2 = 2.10 (.37) *t* (18) = -2.34, *p* = .03	M2 = 2.10 (.37) M3 = 2.13 (.35) *t* (18) = -.70, *p* = .49	M1 = 1.94 (.37) M3 = 2.13 (.35) *t* (16) = -2.31, *p* = .03
Control group	M1 = 1.84 (.34) M2 = 1.82 (.30) *t* (16) = .27, *p* = .79	M2 = 1.82 (.30) M3 = 1.91 (.23) *t* (16) = -1.58, *p* = .13	M1 = 1.84 (.34) M3 = 1.91 (.23) *t* (16) = -.98, *p* = .34
All group	M1 = 1.87 (.31) M2 = 1.90 (.37) *t* (51) = -.55, *p* = .58	M2 = 1.90 (.37) M3 = 1.94 (.31) *t* (51) = -1.44, *p* = .16	M1 = 1.87 (.31) M3 = 1.94 (.31) *t* (51) = -.55, *p* = .58
Student data			
Experimental one		M2 = 3.28 (.48) M3 = 3.31 (.55) *t* (435) = -1.09, *p* = .28	
Experimental two		M2 = 3.34 (.48) M3 = 3.41 (.56) *t* (494) = -2.51, *p* = .01	
Control group		M2 = 3.26 (.45) M3 = 3.26 (.56) *t* (332) = .06, *p* =.95	
All group		M2 = 3.30 (.47) M3 = 3.33 (.56) *t* (1263) = -2.04, *p* =.04	

Notes. M = measurement. The numbers in each column represent the mean score, the standards deviation and the result of the t-test

Appendix 4.2

The Differences in Student Motivation among and between the Three Groups from Measurement to Measurement (*t*-test)

Group	M1 to M2	M2 to M3	M1 to M3
Experimental one	M1 = 2.96 (.44) M2 = 2.93 (.44) *t* (389) = 1.18, *p* = .23	M2 = 2.93 (.44) M3 = 2.90 (.46) *t* (389) = 1.24, *p* = .21	M1 = 2.96 (.44) M3 = 2.90 (.46) *t* (393) = 2.16, *p* = .03
Experimental two	M1 = 3.02 (.44) M2 = 3.00 (.48) *t* (405) = .94, *p* = .35	M2 = 3.00 (.48) M3 = 2.94 (.43) *t* (406) = 2.44 *p* = .01	M1 = 3.02 (.44) M3 = 2.94 (.43) *t* (406) = 2.88, *p* = .00
Control group	M1 = 2.85 (.44) M2 = 2.93 (.43) *t* (326) = -2.21, *p* = .03	M2 = 2.93 (.43) M3 = 2.95 (.47) *t* (325) = -.88, *p* = .38	M1 = 2.85 (.44) M3 = 2.95 (.47) *t* (325) = -3.10, *p* = .00
All group	M1 = 2.95 (.45) M2 = 2.95 (.45) *t* (1123) = -.02 , *p* = .98	M2 = 2.95 (.45) M3 = 2.93 (.45) *t* (1122) = 1.37 , *p* = .17	M1 =2.95 (.45) M3 = 2.93 (.45) *t* (1126) = 1.24, *p* = .21

Notes. M = measurement. The number in each column represents mean score, standards deviation and the result of t-test

Appendix 4.3

The Differences in Teaching Quality Based on School and Teacher Characteristics

Characteristics	Mean score (SD)	*t* /*F* test
School		
Province		*t* (50) = -.68, *p* = .49
1. DKI Jakarta	1.87 (.31)	
2. Banten	1.92 (.25)	
School Size		*F* = (2, 2.14), *p* = .12
1. Small	1.92 (.08)	
2. Medium	1.92 (.28)	
3. Big	1.62 (.12)	
School Accreditation		*t* (50) = .75, *p* = .23
1. A	1.94 (.30)	
2. B	1.88 (.27)	
School Score on English nat. exam		*F* (2, 1.13), *p* = .33
1. Low	1.80 (.23)	
2. Medium	1.92 (.30)	
3. High	1.99 (.22)	
Teacher		
Gender		*t* (50) = 1.397, *p* = .16
1. Female	1.94 (.28)	
2. Male	1.83 (.27)	
Degree		*F* (2, .15), *p* = .86
1. Diploma	1.86 (.24)	
2. Bachelor	1.90 (.29)	
3. Master	1.95 (.24)	
Major		*t* (48) = .13, *p* = .89
1. English	1.89 (.28)	
2. Non-English	1.88 (.25)	
Age		*F* (2, 3.07), *p* = .06
1. <= 30 years	1.88 (.26)	
2. 6 - 10 years	2.02 (.30)	
3. > 10 years	1.75 (.25)	
Years of teaching experience		*F* (2, .09), *p* = .91
1. <= 5 years	1.91 (.28)	
2. 6 - 10 years	1.90 (.29)	
3. > 10 years	1.87 (.27)	

Appendix 4.4

The Differences in Student Outcome Based on School, Teacher, and Student Characteristics

Characteristics	Student Achievement (mean score, SD, *t*/F test)	Student Motivation (mean score, SD, t/F test)
School		
Province	***t* (1658) = 7.52, *p* = .00**	***t* (1118) = 1.85, *p* = .06**
1. DKI Jakarta	9.99 (2.50)	2.97 (.32)
2. Banten	9.06 (2.56)	2.93 (.31)
School Size	***F* (2, 9.28), *p* = .00**	***F* (2, 3.85), *p* = .02**
1. Small	8.23 (2.72)	2.85 (.30)
2. Medium	9.57 (2.60)	2.96 (.31)
3. Big	9.52 (2.03)	2.90 (.35)
School Accreditation	***t* (1617) = 7.47, *p* = .00**	***t* (1118) = .39, *p* = .69**
1. A	10.26 (2.51)	2.95 (.29)
2. B	9.21 (2.55)	2.94 (.32)
School Score on English nat. exam	***F* (2, 28.30), *p* = .00**	***F* (2, 9.57), *p* = .00**
1. Low	8.87 (2.60)	2.89 (.30)
2. Medium	9.53 (2.56)	2.95 (.31)
3. High	10.68 (2.14)	3.05 (.35)
Teacher		
Gender	***t* (1658) = .11, *p* = .91**	***t* (1118) = - .15, *p* = .18**
1. Female	9.52 (2.56)	2.95 (.30)
2. Male	9.51 (2.58)	2.94 (.32)
Degree	***F* (2, 9.950), *p* = .00**	***F* (2, 9.23), *p* = .00**
1. Diploma	8.91 (2.41)	2.86 (.28)
2. Bachelor	9.61 (2.63)	2.95 (.33)
3. Master	9.93 (2.05)	3.02 (.25)
Major	***t* (1606) = 5.41, *p* = .00**	***t* (1118) = -1.63, *p* = .10**
1. English	9.68 (2.61)	2.94 (.31)
2. Non-English	8.72 (2.22)	2.98 (.32)
Age	***F* (2, 8.98), *p* = .00**	***F* (2, 3.99), *p* = .02**
1. <= 30 years	9.23 (2.64)	2.92 (.31)
2. 6 - 10 years	9.85 (2.48)	2.97 (.33)
3. > 10 years	9.55 (2.49)	2.97 (.31)

Years of teaching experience	***F* (2, 17.74), *p* = .00**	***F* (2, 10.24), *p* = .00**
1. <= 5 years	9.24 (2.38)	2.93 (.31)
2. 6 - 10 years	9.14 (2.55)	2.89 (.33)
3. > 10 years	10.05 (2.76)	3.00 (.30)
Student		
Gender	***t* (1658) = -4.48, *p* = .00**	***t* (1118) = -2.98, *p* = .00**
1. Female	9.79 (2.58)	2.97 (.32)
2. Male	9.22 (2.53)	2.92 (.32)
Father education	***F* (3, 21.75), *p* = .00**	***F* (3, 3.90), *p* = .01**
1. Primary	9.02 (2.60)	2.91 (.32)
2. JSS	9.35 (2.46)	2.92 (.32)
3. SSS	9.98 (2.50)	2.99 (.30)
4. University	10.98 (2.50)	2.95 (.31)
Mother education	***F* (3, 18.31), *p* = .00**	***F* (3, 3.40), *p* = .02**
1. Primary	9.13 (2.51)	2.91 (.34)
2. JSS	9.60 (2.64)	2.97 (.30)
3. SSS	10.04 (2.38)	2.98 (.30)
4. University	11.04 (2.67)	2.89 (.30)
Father job	***F* (2, 26.86), *p* = .00**	***F* (2, .23), p = .79**
1. Labor and farmer	9.03 (2.43)	2.95 (.32)
2. Small business	9.67 (2.60)	2.94 (.32)
3. Professional	10.36 (2.49)	2.96 (.29)
Mother job	***F* (3, 6.22), *p* =.00**	***F* (3, 2.12), *p* = .09**
1. Housewife	9.56 (2.54)	2.95 (.31)
2. Labor and farmer	8.74 (2.50)	2.94 (.29)
3. Small business	9.35 (2.61)	2.86 (.35)
4. Professional	10.44 (2.54)	2.99 (.33)

APPENDICES FOR CHAPTER 5

Appendix 5.1

The Results of the Multilevel Analysis of the Mediating Role of Teaching Improvement on the Covariance Analysis

	Model 5 (TI)		Model 6 (Intervention and TI)		Model 7 (Interaction between intervention and TI	
	Coeff	SE	Coeff	SE	Coeff	SE
Fixed Part						
Intercept	8,206***	.563	7,389***	.609	7.490***	.619
Student level						
Pretest	.146***	.028	.146***	.028	.146***	.028
Gender (female)	.392***	.155	.397***	.155	.400***	.155
Father education (JHS)	-0.211	.216	.223	.216	-.227	.216
Father education (SHS)	-0.116	.212	-.136	.212	-.146	.212
Father education (univ)	0.554	.339	.510	.339	.489	.339
School level						
Province (Banten)	-2,048***	.478	-1,657***	.462	-1.698***	.451
School nat. exam (med.)	.704	.536	.516	.503	.516	.495
School nat. exam (high)	2,804***	.971	1,905*	.952	2.408*	.979
Intervention						
Intervention one			.905	.545	.826	.544
Intervention two			1.668**	.559	1.665**	.593
Teaching improvement (TI)	.630	.729	.051	.722	-.670	1.329
Interaction effect						
Intervention one and TI					2.314	1.773
Intervention two and TI					.108	1.66
Random Part						
School level variance						
Intercept	2,389	.524	2.003	.450	1.890	.428
Student level variance						
Intercept	7,317	.292	7,316	.292	7.316	.292
Deviance (-2*loglikelihood)	6,454,486		6,446,216		6443.585	
Decrease in deviance	.744		.500		2.636	
Variance explained	0.00		.00		.01	

*p < .05 (2-tailed), ** p < .01 (1-tailed), *** p < .01 (2-tailed)

Appendix 5.2

The Results of the Multilevel Analysis of the Mediating Role of Teaching Improvement on Learning Gain

	Model 5 (TI)		Model 6 (Intervention and TI)		Model 7 (Interaction between intervention and TI	
	Coeff	SE	Coeff	SE	Coeff	SE
Fixed Part						
Intercept	1.045*	.527	.552	.596	.638	.618
Student level						
Gender (female)	.128	.203	.130	.203	.133	.203
Father education (JHS)	-.416	.284	-.437	.284	-.442	.284
Father education (SHS)	-.519	.277	-.556	.277	-.569	.277
Father education (univ)	-.386	.441	-.459	.442	-.487	.442
School level						
Province (Banten)	-1.715***	.472	-1.455***	.473	-1.481	.470
School nat. exam (med)	-.223	.529	-.326	.516	-.333	.516
School nat. exam (high)	1.182	.954	.606	.972	.920	1.016
Intervention						
Intervention one			.422	.557	.353	.565
Intervention two			1.154**	.572	1.115	.618
Teaching improvement (TI)	.834	0.727	.374	.745	-.232	1.382
Interaction effect						
Intervention one and TI					1.574	1.842
Intervention two and TI					.318	1.741
Random Part						
School level variance						
Intercept	2.066	.509	1.875	.472	1.830	.463
Student level variance						
Intercept	12.684	.505	12.684	.505	12.685	.505
Deviance (-2*loglikelihood)	7145.806		7141.759		7140.816	
Decrease in deviance	1.305		.253			
Variance explained	.00		.00		.00	

*$p < .05$ (2-tailed), ** $p < .01$ (1-tailed), *** $p < .01$ (2-tailed)

Appendix 5.3

The Results of the Multilevel Analysis of the Mediating Role of Teaching Improvement on Student Motivation

	Model 7 (TI)		**Model 8 (Intervention, interaction between time and intervention and TI)**		**Model 9 (Intervention, interaction between time and intervention, TI, and interaction between TI and intervention)**	
	Coeff	SE	Coeff	SE	Coeff	SE
Fixed Part						
Intercept	2.754****	.052	2.657****	.056	2.632****	.056
Time	-.012	.011	.048*	.019	.049	.019
Student level						
Gender (female)	.062***	.018	.060***	.018	.060***	.018
Teacher level						
Teacher degree (bachelor)	.102	.055	.104	.052	.107	.050
Teacher degree (master)	.230****	.082	.223****	.077	.210****	.073
School level						
School nat. exam (medium)	.068	.050	.050	.048	.107	.050
School nat. exam (high)	.276****	.081	.221****	.079	.210****	.073
Intervention						
Intervention one			.129	.057	.148	.055
Intervention two			.201***	.056	.244***	.057
Interaction effect						
Time X intervention one			-.079***	.025	-.079***	.025
Time X intervention two			-.092***	.025	-.093***	.025
Teaching improvement	.098	0.058	.071	.057	.272	.106
Interaction effect						
Intervention one and TI					-.195	.146
Intervention two and TI					-.314	.130
Random Part						
School level variance						
Intercept	.007	.003	.005	.003	.004	.003
Time	.002	.001	.002	.001	.002	.001
Student level variance						
Intercept	.041	.004	.041	.004	.041	.004
Time level variance						
Intercept	.145	.004	.144	.004	.144	.004
Deviance (-2*loglikelihood)	3759.673		3741.671		3736.072	
Decrease in deviance	2.762		1.500		7.099	
Variance explained	.01		.01		.01	

* $p < .05$ (1-tailed), ** $p < .05$ (2-tailed), *** $p < .01$ (1-tailed), **** $p < 0.01$ (2- tailed)

Siti Nurul Azkiyah

The Effects Of Two Interventions

ON TEACHING QUALITY AND STUDENT OUTCOME